Exploring the sexual phase of the diatom -nitzschia multistriata</i: from genes to metabolites

Thesis

Exploring the sexual phase of the diatom *Pseudo-nitzschia multistriata*: from genes to metabolites

Camilla Borgonuovo, MCs

Thesis submitted for the degree of

Doctor in Philosophy (PhD)

School of Life, Health and Chemical Sciences

Open University

Stazione Zoologica Anton Dohrn

Director of Studies: Dr Maria Immacolata Ferrante

Internal Supervisor: Dr Marina Montresor

External Supervisor: Prof Wim Vyverman

December 2018

Abstract

Phytoplankton are an important component of the marine ecosystems since they are at the base of marine food webs and are responsible for about 50% of the oxygen production at a global scale. The knowledge of phytoplankton life cycles is important to understand the mechanisms regulating the dynamics of their populations in the natural environment but also to improve the capability to cultivate them and thus exploit their biotechnological potential. The model organisms chosen for my PhD project is a marine planktonic diatom, *Pseudo-nitzschia multistriata*, known to produce the neuro-toxin domoic acid (DA), the causative agent of Amnesic Shellfish Poisoning.

Pseudo-nitzschia multistriata is a pennate heterothallic species whose life cycle has been described and can be controlled in laboratory conditions. Cells of the opposite Mating Types (MTs) can switch from asexual to sexual reproduction after reaching a threshold size and it is very important that this process is finely synchronized and regulated to ensure its success. There is evidence of a chemical cross talk that mediates mating and some of the genes involved in the process were identified.

In this thesis, several aspects related to the reciprocal perception of the opposite MTs and the comparison a between a 'treatment' in which sexual reproduction was ongoing as compared to the 'control' of monocultures of single parental strains were untangled focusing on a set of genes involved in the process.

Several experiments illustrated in Chapter 2 were aimed at i) further elucidating the expression pathways of the target genes and ii) identifying a set of genes that could be used as marker genes for future isolation of putative pheromones. The results of the experiments suggest that in *P. multistriata* there are two constitutive pheromones called *MRP1* and 7488

produced by MT+ and MT-, respectively. Among these the gene 7488 was upregulated in MT- exposed to the culture medium in which MT+ was growing; this gene can thus be considered the as the molecular marker for a bio-assay to detect the fraction of MT+ conditioned medium containing the putative pheromone. The experiments carried out in this thesis also suggested that that the constitutive putative pheromone of MT+ perceived by MT- could be MRP1, a gene encoding for a small secreted protein, that is differentially expressed between opposite MTs and is highly induced during sex in the MT+.

In Chapter 3, I have illustrated the spatial pattern of a set of genes involved in the sexual phase using the large database of genes from the Tara Oceans expedition. The co-occurrence of MT related genes and one meiotic gene in nine TARA stations suggests that sexual reproduction was occurring in those sites. The spatial distribution of these genes was not uniform and this can be explained with differences in their basal expression levels and/or by the absence of species in which the genes are present.

Finally, in Chapter 4 I explored the difference between metabolites produced by the parental strains in vegetative growth in monoculture and by a co-culture of strains of opposite MT undergoing sex. The metabolomics analysis revealed that there were not exclusive metabolites distinguishing the vegetative phase and the sexual phase, but there were a number of mostly unidentified metabolites that increased their quantity in the sexual phase. The results of this thesis add novel information on several aspects concerning the mate perception of a planktonic diatom, elucidating hierarchical activation of genes during the sexual phase. Furthermore, with future identification of the constitutive cue of MT+, MT- gene expression changes in response to the first MT+ signal might be clearly elucidated using RNAseq. Last but not least, this work indicates that genes related to sex could be makers for sexual detection at sea, becoming a useful tool for ecological purposes.

Index

List of figures

List of tables

Chapter I:

General introduction

1.1 Diatoms

Diatoms are photosynthetic unicellular algae assigned to the *Bacillariophyceae*, a class of the *Heterokonta* phylum (also called *Stramenopiles*) that includes diatoms, brown macroalgae and plant parasites (a well-known example is *Phytophthora*, the agent causing the Irish potato famine) (Round, Crawford, and Mann 1990). They are widely distributed in both marine and freshwater ecosystems, and can also live in the moist soil. About one-fifth of the photosynthesis on Earth is carried out by diatoms. Diatoms can be benthic or planktonic and, depending on their habitat, they have distinctive morphological characteristics; they can form colonies in chains. Diatom cell size can span from 5 to 200 µm and they have a characteristic cell wall made of silica. Sediments made of diatom frustules (diatomites) are used for several commercial applications.

1.2 Morphology and diversity

The word 'diatom' derives from the word Greek "διατομος" that means «cut in half». This word is referred to a morphological feature that is their two-part cell wall made of silica, called frustule. The frustule consists of two slightly unequal parts, like a box, the hypo-theca, with its lid, the epi-theca. Both epi-theca and hypo-theca are constituted of several parts: the valve, similar to a plate, and one or several cingular bands (Benton and Harper 2009) (Figure 1.1). The frustule is essentially made of hydrated glass ($SiO_2.nH_2O$). Diatoms are fascinating example of how nature is art even in the microscopic world. Indeed, their frustule shows several shapes and decorations making them glass bling. Diatoms are the most diverse group of phytoplankton with ca. 100,000 species estimated (Mann and Vanormelingen 2013). The diversity is overshadowed by the cryptic species, that are species with similar morphological phenotype but genetically different. Examples of cryptic species are reported for the *Skeletonema* and *Pseudo-nitzschia* genera (Orsini et al. 2004; Amato et al. 2007). The estimated diversity has increased considerably thanks to

recent studies that generated a dataset of 4,748 OTUs for diatoms species using metagenomics data of the global TARA expedition (Malviya et al. 2016).

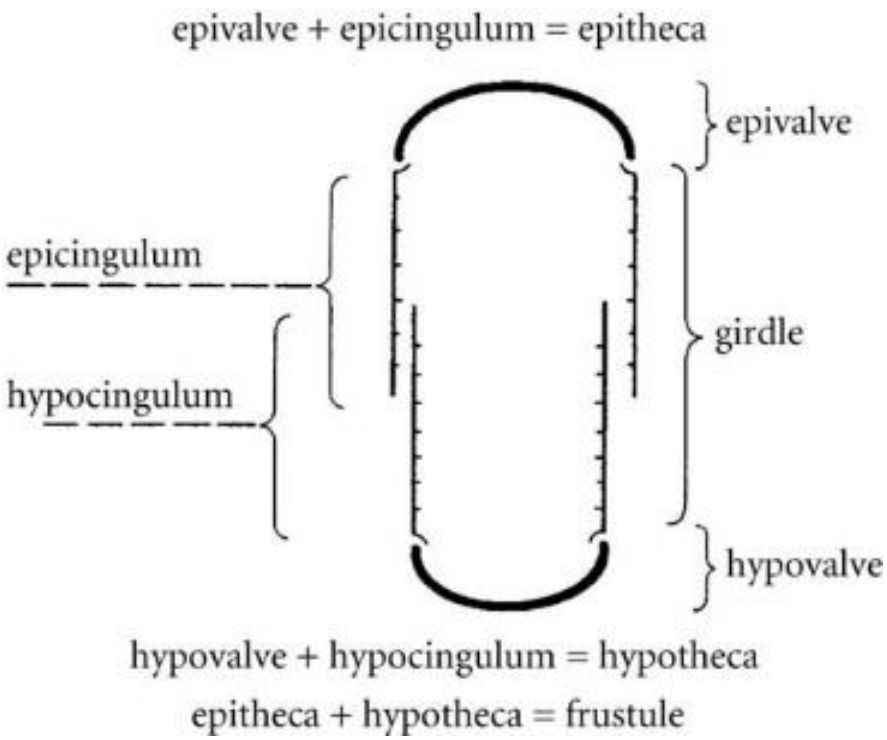


Fig.1.1 Schematic representation of diatom morphology. From "Introduction to Paleobiology and the Fossil Record", Benton, M., & Harper, D. A.

1.3 Ecology

Global diatom primary production is considered to be equivalent to all the combined primary production of tropical rainforests (Field 1998). Furthermore, they have an important role for aquatic food webs and consequently also for fisheries. Diatoms control the biogeochemical cycle of silicon in the oceans by incorporating dissolved silicic acid into their frustule. This silicon can then be sequestered on the ocean floor in the form of accumulation of frustules of dead diatoms, or can return to dissolved silicic acid in sea water following dissolution of the thinner frustules (Treguer et al. 1995). The sinking time of frustules is slower than the incorporation of silicon in the new frustule, silicon passes through the biological uptake and dissolution cycle an average of about 39 times before being removed to the seabed (Treguer et al. 1995). The importance of diatoms is also

related to the carbon cycle, sequestering it into the deep ocean for thousands to millions of years (Falkowski 1998). Diatoms are very abundant in costal and upwelling zones rich in nitrate (NO_3^-), while they are less abundant in the open ocean where nitrate concentration is low (V. Smetacek 1999). Diatoms can store nitrate and they can affect also the nitrogen cycling because they sink at the end of their bloom period. Nutrient depletion can induce mass sinking of diatoms (V. S. Smetacek 1985). The bio-available iron concentration in the sea is often limiting for diatoms. Experiments of iron fertilization in high nutrient low chlorophyll oceanic regions (HNLCs) highlighted that iron has an important role for diatoms, for example in the HNLC of Southern Ocean iron is the only limiting factor for diatom and phytoplankton in general growth (Watson et al. 2000). Lastly, some species of diatoms can form harmful blooms: species of the genus *Pseudo-nitzschia* produce domoic acid, a neurotoxin that bioaccumulates in bivalves. The first food poisoning caused by domoic acid (Amnesic Shellfish Poisoning) was reported in Prince Edward Island (Canada) in November 1987 (Bates et al. 1989).

1.4 Evolutionary history

Diatoms originated about 250 million years ago. They derived from two endosymbiotic events; in the primary event of endosymbiosis, that happened about 1.5 billion years ago, a cyanobacterium was engulfed by a eukaryotic heterotroph and this event originated the photosynthetic plastids of the plants, red and green algae. Independently, about 500 million years later, an eukaryotic heterotroph incorporated a red alga (the secondary event of endosymbiosis) to generate the photosynthetic plastids of diatoms, brown algae and plant parasites(Yoon et al. 2004; Armbrust 2009) (Figure 1.2).

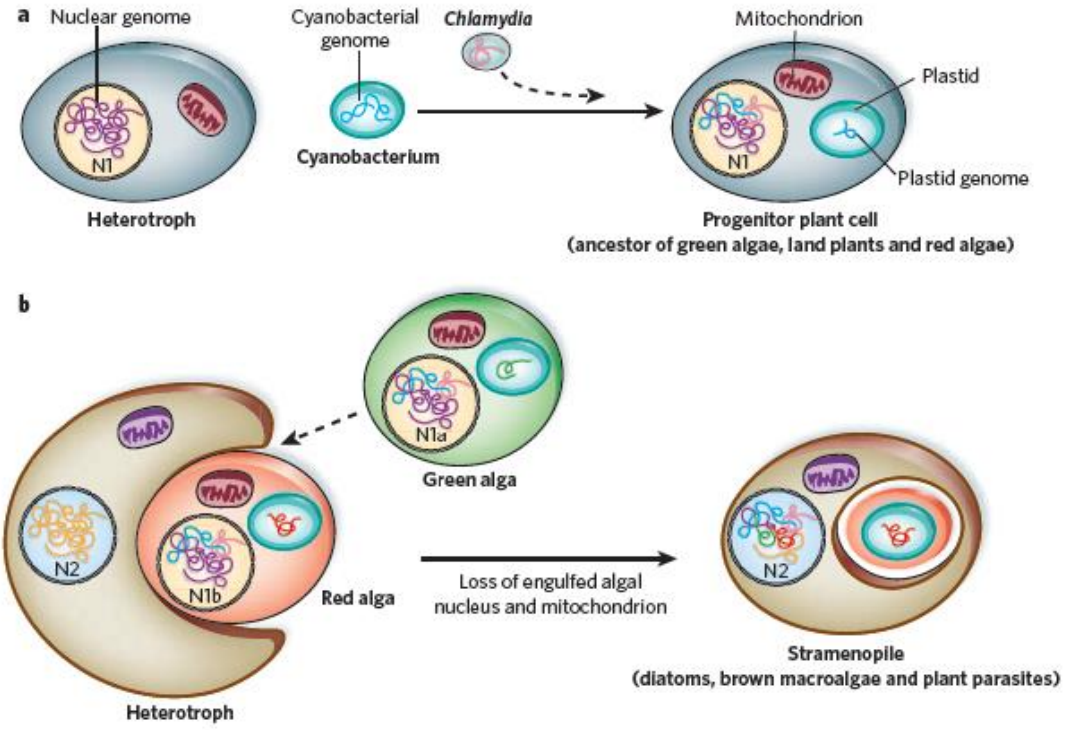


Fig.1.2. a. The origin of the plant progenitor plastid (primary endosymbiosis); b. the origin of Stramenopile plastid (secondary endosymbiosis). Cartoon from Armbrust (2009).

Diatoms are broadly divided into two main groups, centrics and pennates, depending on their shape; centric diatoms, mainly planktonic, have circular valves while pennate diatoms, mainly benthic, have elongated valves. Bipolar and multipolar centric diatoms

originated about 150 million years ago while pennate diatoms evolved about 96 million years ago (Sorhannus 2007). Pennates are further divided into araphid and raphid. The oldest araphid pennates arose around 93 million years ago, while the more recent raphid pennate diatoms evolved about 75 million years ago (Sorhannus 2007). The valves of raphid pennates have a raphe, a slit in the valve from which polysaccharides are secreted and allow cells to slide on the substrate or to adhere on it. The raphe allowed diatoms to colonize a wide range of new benthic habitats (Sims, Mann, and Medlin 2006).

1.5 Diatom life cycles

Diatoms have a diplontic life cycle, which includes the diploid phase that is the vegetative cell and the haploid phase that is represented by gametes. The life cycle of centric and pennate diatoms is illustrated in Figure 1.3 (Montresor et al. 2016). Centrics and pennates produce gametes with different morphology; the gametes of centrics have unequal size and morphology with large egg cells and small and motile sperm cells (anisogametes), while the gametes of pennates are morphologically similar (isogametes). This difference in gamete morphology could be related to different habitats that centrics and pennates colonize. Centric diatoms have flagellate male gametes that can reach the female egg cell in the water column, while most of the pennates have non-motile gametes, that can fuse when male and female gametangia slide close to each other on the substrate. *Pseudostaurosira trainorii,* an araphid pennate, produces a motile male gametes with structure similar to flagella called threads (Sato et al. 2011).Furthermore, the centrics are mainly homothallic while the pennates are mostly heterothallic. The same monoclonal culture of centrics can produce both egg and sperm cells, depending on the cell size of them in the culture, while the pennates have two separated sexes, called mating types (MT), and only if two opposite MTs (MT + and MT-) are in close contact the gametes are produced (Montresor et al. 2016). Atypical cases like autogamy (fusion of two nuclei of the same gametangium), paedogamy (fusion of two gametes of the same gamentangium),

apomixes (fusion of pseudo-gametes produced by pseudo-meiosis), and parthenogenesis (development of an unfused gamete into an zygote that produces a clonal lineage) were reported (Kaczmarska et al. 2013). In this complex framework of the life cycle, diatoms face a critical moment during their lifetime, that is the reduction of cell size. The cell size reduction is caused by the particular morphology of the frustule formed by two unequal valves assembled like a box with its lid. With vegetative division each of the two daughter cells form a new hypotheca inside the theca inherited by the mother cell. The consequence of this cell division modality is that one of the daughter cells will be smaller than the mother cell, with consequent and progressive reduction of cell size in the population. It is clear that this will lead to extinction of the population without a mechanism to escape the miniaturization process. In most diatoms, the solution to this problem is sexual reproduction. Diatoms cannot undergo sexual reproduction over the whole cell size range, but cells become competent for sex only when they reach a species-specific size threshold (SST). Then thanks to the sexual phase the population restores the maximum cell size of the species.

The mechanism to restore cell size has not been observed in all diatoms; in some cases, vegetative enlargement can occur. Moreover, not all diatoms experience cell reduction, for example *Thalassiosira pseudonana* and *Phaeodactylum tricornutum* do not reduce their size (Montresor et al. 2016).

The triggers of sexual reproduction are generally related to environmental factors, mostly in centrics, and to perception of the opposite MT, in the heterothallic pennates.

Detailed information about the life cycle of diatoms are mainly available for two species: the benthic pennate *Seminavis robusta* (Chepurnov et al. 2002) and the planktonic pennate *Pseudo-nitzschia multistriata* (Scalco et al. 2016).

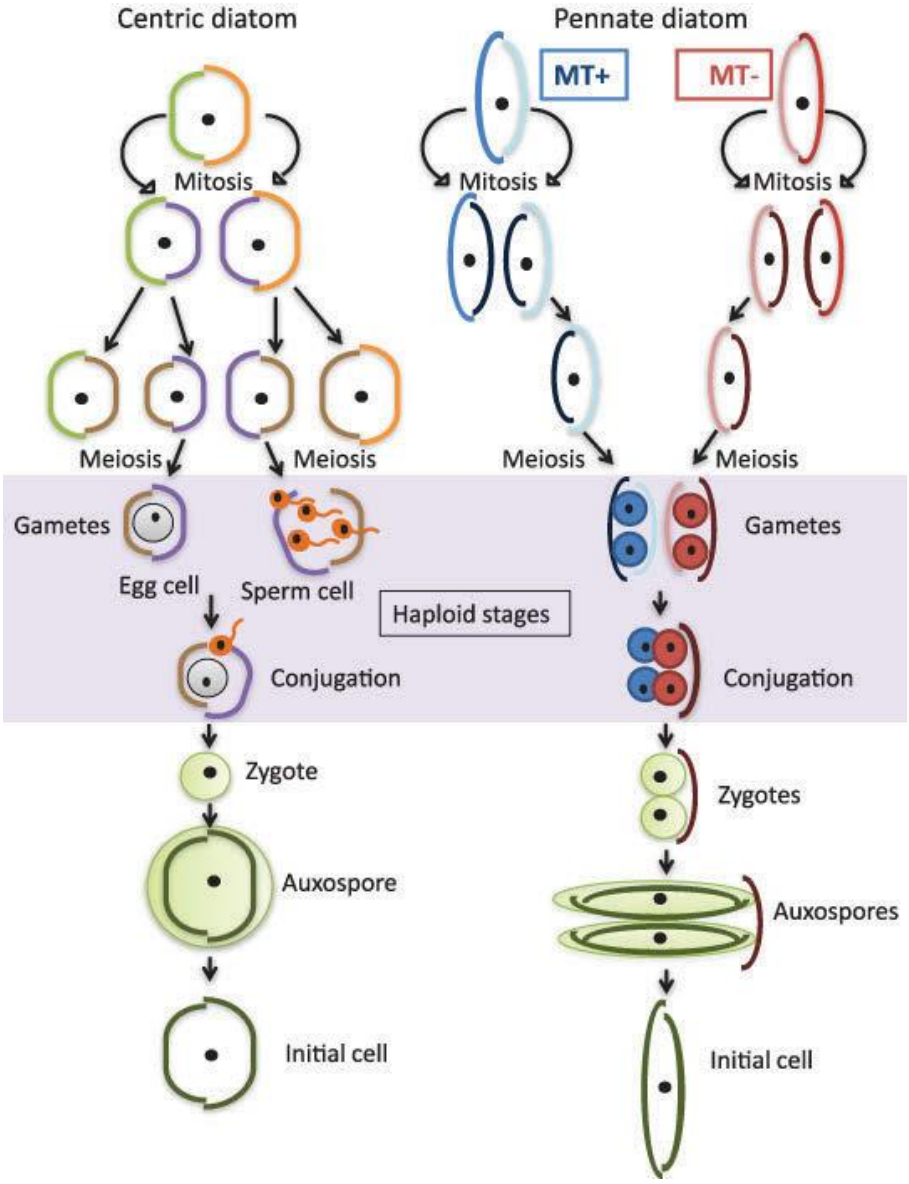


Fig.1.3 Schematic drawing of the life cycle of a centric and a pennate diatom. Diatom cells are diploid and are surrounded by a rigid frustule made of two unequal thecae. During mitosis, the new thecae are formed inside the frustule. This causes a progressive decrease in the population cell size. The formation of gametes takes place following meiosis in cells (gametangia) that are below a species-specific size threshold for sexualization. In centric diatoms, large macrogametes (egg cells) and small uniflagellated microgametes (sperm cells) are produced within the same strain. In pennate diatoms, the formation of gametes occurs when two strains of opposite mating type are in close contact; gametangial cells pair side to side and meiosis takes place. Conjugation of the haploid gametes produces a zygote that expands into an auxospore. Within the auxospore, the large initial cell is synthesized (Montresor et al. 2016).

1.6 Life cycle of *Pseudo-nitzschia multistriata*

Pseudo-nitzschia multistriata is a pennate heterothallic species, that's there are two separate MT. The cells are not able to undergo sexual reproduction until they reach the SST, 55 µm in this species (D'Alelio et al. 2010). The sexual phase occurs when two opposite and competent MTs meet. During the early phase of mating the cells should be able perceive the opposite MT to allow the formation of pairs of two gametangia. After pair formation, meiosis and gamete formation occur. In each gametangium two gametes are produced following meiosis; in each gamete there are two chloroplasts and two nuclei, of which one is degraded. After the formation of gametes, the frustules of gametangia open up to release gametes, each with one nucleus and two chloroplasts. Each gamete fuses with the gamete of opposite MT to form the zygote that remains attached to the frustule of the mother cell. The zygote, with four chloroplasts and two haploid nuclei, grows to produce the auxospore. The elongation of the auxospore is completed after about 20 hours and the new frustule is deposited inside the auxospore forming the new cell, called initial cell, which is about 80 µm long. The initial cell with one diploid nucleus and four chloroplasts exits out from perizonium, during the first division, in which chloroplasts do not duplicate, and finally two normal vegetative diploid cells with two chloroplasts are produced (Scalco et al. 2016). In Figure 1.4 the life cycle of *Pseudo-nitzschia multistriata* is shown (Basu et al. 2017)

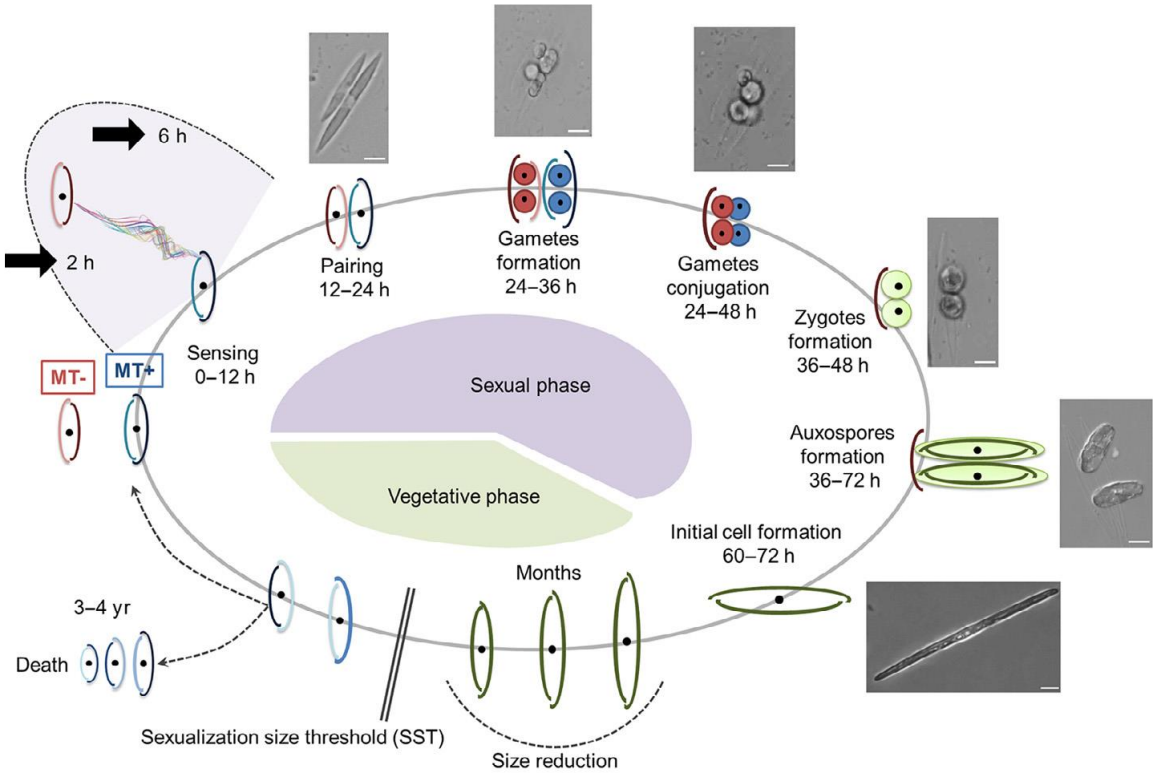


Fig.1.4 Schematic drawing of the life cycle of *Pseudo-nitzschia multistriata*. Starting clockwise from the bottom portion of the cycle, the vegetative phase is characterized by progressive cell size reduction of the population imposed by the rigid silica wall, made up of two unequal thecae. During this process, the cells reach the sexualization size threshold (SST) and can either keep decreasing in size until death, or undergo sexual reproduction and escape the miniaturization process, producing large cells. In the heterothallic *P. multistriata*, sex can occur only if strains of opposite mating type come into contact. The perception of chemical cues deriving from the mating partner (0–12 hours) brings cells of opposite mating type to pair (12–24 hours). The formation of gametes (24–36 hours) takes place following meiosis. Conjugation of the haploid gametes (24–48 h) produces two expandable zygotes (36–48 h) that develop into auxospores (36–72 hours). Within each auxospore, an initial cell of maximum size is synthesized (60–72 hours), restoring the vegetative phase of the cycle. The time interval for each stage is indicated. Representative microscopic images of the different stages are shown outside the circle; bar, 10 μm. MT, mating type (Basu et al. 2017).

1.7 Sex pheromones in algae

The study of chemical cues is important to understand population dynamics and diversity of phytoplankton communities since they mediate interactions in and among populations. Usually, the interaction among organisms of the same species is mediated by pheromones while interspecific communication is mediated by allelochemicals (Sabelis and Dicke 1987). Moreover, in the biotechnology field there is a growing interest for natural compounds produced by phytoplankton that is leading forward the interest in understanding the production and action mechanisms of chemical cues. Among natural compounds, the algae pheromones are the less studied. For organisms that experience sexual reproduction during their life cycle, mating is a crucial moment for successful reproduction. Mating is mediated by sex pheromones that is different in different species; indeed, these chemical signals allow to recognize a sexually mature partner with opposite sex and belonging to the same species. Terrestrial organisms detect sex pheromones in the air, while in aquatic organisms, the signalling molecules should be soluble in water. Within aquatic organisms, few studies are focused on algae. In the unicellular green algae *Chlamydomonas reinhardtii*, gametogenesis is trigged by nitrogen starvation (Starr, Marner, and Jaenicke 1995), and the mating depends on random movement of compatible cells. In *Chlamydomonas allensworthii* an attractive phenotype (TSUBO 1961; I Maier 1993) mediated by pheromones was observed: the attraction pheromone lurlenic acid is produced by the MT- gametes and attracts the motile MT+ gametes (Mori and Takanashi 1996a, 1996b); attraction of gametes from strains that synthetize the different pheromone variants (lurlenon) does not occur. A phylogenetic analysis using the ITS (Internal Transcribed Spacer) 1 and 2 regions in these different strains showed at least two cryptic lineages within this species reflecting the presence of two pheromone response types (Coleman, Jaenicke, and Starr 2001). In *Volvox* heat shock induces the production of a sex inducer glycoprotein by the male, this sex inducer acts on male and female vegetative colonies inducing gametes production in both, and then male gametes produce the same

compound that turns into pheromone to synchronize the mating event (Frenkel, Vyverman, and Pohnert 2014). Sex inducer compounds have been observed in several species of *Volvox* but their chemical nature is different, for example in *V. capensis* L-glutamic acid is employed (Frenkel, Vyverman, and Pohnert 2014). In brown algae, the first evidence of the production of sex pheromones came from *Ectocarpus siliculosus* (Müller et al. 1971) where the diploid sporophyte produces zoospores, which develop into the haploid male and female gametophytes that produce motile gametes (Coelho et al. 2012). In this species, the male-attracting substance was characterized as ectocarpene that is a fatty acid derived hydrocarbon (Müller et al. 1971). In *Laminaria digitata*, the release of male gametes is synchronized by an epoxidized hydrocarbon that is structurally related to ectocarpene (Ingo Maier and Müller 1982). Ectocarpene and similar compounds were also found in diatoms. Here, however, the function of these compounds was not related to sexual reproduction but to chemical defence (Pohnert, 2002). In pennate diatoms, sexual reproduction involves many signalling processes between opposite MTs below a sexual size threshold (SST). The first evidence of pheromones in araphid pennate diatoms came from the heterothallic *Pseudostaurosira trainorii* (Sato et al. 2011), a heterothallic and anisogamous species with motile MT+ gametes that can move by threads (structures like flagella) to fertilize the non-motile MT- gamete. In *Pseudostaurosira trainorii*, sexual reproduction occurs with the action of three putative pheromones: a first MT- sex pheromone secreted by vegetative cells induces sexualisation of MT+ vegetative cells that produce gametes and secrete a second pheromone, this second pheromone stimulates sexualisation in the MT- that produces gametes, and finally a putative third pheromone that is an attractant for MT+ gametes. The motile MT+ gametes move first in a random way and then direct toward the MT- (thanks to the putative third pheromone produced by MT-) get in contact and eventually fertilize the not motile MT- gametes. In this species sexual reproduction occurs in light as well as in dark condition (Sato et al. 2011). The existence of these pheromones was demonstrated by behavioural experiments, but their chemical nature

is unknown, while the first sex pheromone chemically characterized in diatoms was the small amino acid diproline in the benthic pennate *Seminavis robusta* (Gillard et al. 2013). Diproline is produced by MT- only during the mating event process and furthermore the attraction behaviour was observed only when diproline was detectable in the medium; these observations demonstrated that the diproline is an attractive pheromone. The attraction is a fundamental step for gametangia to ensure pairing and then the fusion of gametes (not motile in this species). The production of diproline starts after illumination and when the vegetative cells are in the G1 phase of the cell cycle (Gillard et al. 2008; Gillard et al. 2013). Other two sex pheromones are involved in the sexual reproduction of *S. robusta*: sex-inducing pheromone plus (SIP+) and minus (SIP-). SIP+ was only partially characterized and its structure contained at least one sulphur atom (Moeys et al. 2016). SIP+ induces the arrest of the cell cycle in the G1 phase in MT- cells and, at the same time, it induces the production of diproline. The role of SIP- is yet to understand. In this multi-steps recognition system, receptors and many metabolic pathways are employed that are yet to be explored to better understand the whole phenomenon. *S. robusta* is a benthic species and is not the ideal model to study mate perception in planktonic species.

There is evidence of sex pheromones also in the planktonic species *P. multistriata*, indeed when opposite MTs are in co-culture where only chemical communication is allowed, in both MTs there are gene expression changes, indicating that mate perception occurred even if physical contact was impeded (Basu et al. 2017).

1.8 Available resources to study diatoms

Functional genomic studies require genomic resources like genome sequences and molecular tools like genetic transformation. The first diatom genome sequenced was that of the centric species *Thalassiosira pseudonana* (Armbrust et al. 2004), followed by that of the pennate species *Phaeodactylum tricornutum* (Bowler et al. 2008). Other diatoms sequenced genome are those of *Thalassiosira oceanica* (Lommer et al. 2012), *Fragilariopsis cylindrus* (Mock et al. 2017), *Fistulifera solaris* (Tanaka et al. 2015), *Cyclotella criptica* (Traller et al. 2016) and *P. multistriata* (Basu et al. 2017). Furthermore, about 650 transcriptomes of marine micro eukaryotes, including diatoms, have been generated in the context of the Marine Microbial Eukaryotic Transcriptome Sequencing Project (MMETSP) (Keeling et al. 2014). From all these big omics data it emerged that in diatoms there are several gene families that are expanded respect to other eukaryotes: the cyclins family (Huysman et al. 2010), the heat shock factor family (Montsant et al. 2007) and a range of transcription factors (Rayko et al. 2010). Furthermore, annotation of diatom genes is available only for little more than half of all genes. Functional studies (forward and reverse genetics) aim to understand the function of genes. Diatoms, with their adaptability in a changing environment and plasticity in responding to various biotic and abiotic stresses, represent an opportunity to expand our knowledge about functional diversity of the plankton and exploit, by genetic engineering, this diversity and new gene functions for biotechnology. Fortunately, this investigation is possible thanks to several studies where molecular tools where employed in diatoms to understand gene function. The first method used in diatoms for gene function studies was the bombardment method (gene gun) that was applied in *C. cryptica* and *Navicula saprophila* (Dunahay, Jarvis, and Roessler 1995) and subsequently for *P. tricornutum* (Apt, Grossman, and Kroth-Pancic 1996). Other transformation methods more recently developed in *P. tricornutum* are electroporation (Niu et al. 2012; Zhang and Hu 2014) and bacterial conjugation (Karas et al. 2015; Diner et al. 2016). The genetic transformations allow to explore different aspects,

not only gene function by inactivation and overexpression of genes, but also to understand the spatial localization of a protein in the cell, inserting in the plasmid the target gene fused with a fluorescent reporter like the green fluorescence protein (GFP) (P. Kroth 2007). Another tool developed in diatoms is gene silencing (De Riso et al. 2009); this technique employed small RNAs molecules that inhibit the expression of the target gene. More recently in *T. pesudonana* and in *P. tricornutum*, new methods for genome editing were applied: the TALEN (Transcription Activator-Like Effector Nucleases) (Daboussi et al. 2014) and CRISPR (Clustered Regulatory Interspaced Short Palindromic Repeats)/Cas9 (Hopes et al. 2016; Nymark et al. 2016; P. G. Kroth et al. 2018). These methods allow cutting specific sequences of double-strand DNA that is repaired by non-homologous end-joining or homologous recombination. Among all these techniques, the bombardment method was set up in *P. multistriata* (Sabatino et al. 2015), while gene silencing has been setup in the congeneric species *P. arenysensis* (Sabatino PhD thesis, Silencing of lipoxygenase pathways in the diatom genus *Pseudo-nitzschia*).

1.9 *Pseudo-nitzschia multistriata*: state of the art

The genus *Pseudo-nitzschia* is known for domoic acid (DA) production. It was reported that there is variability in DA production among different species but also among different strains of the same species (Bates et al. 1989; Bates, 2018.). The genus forms blooms in coastal waters and open oceans. The morphological characteristics that allow to identify several species of *Pseudo-nitzschia* are cell shape and width, density of striae and fibulae, morphology and density of perforations (Lundholm et al. 2006; Amato and Montresor 2008; Lim et al. 2018). To explore the genetic diversity of this genus, analyses were carried out using several molecular markers like ITS (internal transcribed spacer region between the small-subunit ribosomal RNA and large-subunit ribosomal RNA genes, with high sequence variation) and LSU (large subunit of ribosomal DNA containing variable domains at the 5’ end conserved regions at the 3’ end in the sequence) (Lundholm et al.

2006; Amato and Montresor 2008). *Pseudo-nitzschia multistriata* is recorded at the Long Term Station MareChiara in the Gulf of Naples since 1995 (D'Alelio et al. 2010). It blooms in summer and early autumn and produces the neurotoxin DA (Orsini et al. 2002). Recently, the genes involved in DA production have been discovered in *P. multiseries* (Brunson et al. 2018) and these genes were found with the same genetic organization in the *P. multistriata* genome.

Using 10 years of data from the Gulf of Naples, the life history of *P. multistriata* was determined by analysing abundances and size of cells in the samples collected on a weekly basis. Cell size showed an alternation of uni- and bi-modal distribution over time and the largest size classes were recorded every two years. The presence of large cells every two years, was interpreted as signature of sexual reproduction and a progressive cell size reduction of the cohorts originating from the large initial cells was detected between two events of sexual reproduction (uni- and bi-modal distribution of cell size) (D'Alelio et al. 2010). This biennial pattern was confirmed by a model parametrized by data on cell size reduction, growth rates, and sexualisation size threshold. The model predicts that *P. multistriata* will become locally extinct if sexual reproduction does not occur within 4 years.

In a recent study, hundreds of strains from the Gulf of Naples, covering two consecutive years, were genotyped with 22 microsatellite markers (Ruggiero et al. 2018) and two events were observed. The first was that in 2013 92.6% of strains were MT+, while in 2014 the sex ratio was 1:1; the second was that the samples from the 2013 autumn bloom showed an extremely low genotypic diversity only differing for a few alleles, which is evidence for a clonal expansion event, never recorded up to now for marine diatoms.

In *P. multistriata* sexual reproduction is density dependent, in other words the signal that triggers sexual reproduction is perceived when the distance between opposite MTs is

reduced (Scalco et al. 2014). Another important aspect is the physiological state; indeed successful sexual reproduction occurs when parental strains are in the exponential growth phase (Scalco et al. 2014). In optimal cross conditions it was observed also that the growth of parental cells, in the presence of sexual stages, is significantly reduced respect to vegetative parental cells. In *P. multistriata*, there is no attractive behaviour between opposite MTs, contrary to *S. robusta* (Gillard et al. 2013). A descriptive study provided details on all phases during sexual reproduction, from the random movement of both MTs to the formation of gametangia couples to the formation of initial cells, providing useful reference data for deeper studies on this important phase of the life cycle of planktonic pennate diatoms (Scalco et al. 2016).

Only recently genomics and transcriptomic resources became available for this species (Basu et al. 2017). The sequencing and assembly of a clonal diploid strain provided a genome of 59 Mb with a total of 12,008 genes of which 9,653 were assigned a UNIPROT ID, while 214 genes were annotated only for the presence of a protein domain. In the *P. multistriata* genome 11 families showed expansion within the diatom lineage, 26% of the proteome was predicted to be orphan, 252 genes were predicted to be of bacterial origin and 123 were classified to be of red algal origin. About 1500 conserved non-coding elements were identified mainly localized near transcription start sites, and about 25% of the genome was constituted of repetitive elements with 6% of these being long terminal repeat retrotransposons. RNAseq was performed to highlight transcriptional changes during the early phase of sexual reproduction (two and six hours from the beginning of the experiment) in MT+ and MT-. The experiment was carried out using a glass bipartite apparatus with an insert membrane allowing chemical communication without physical contact between the cells (Figure 1.5). Flow cytometry analysis showed that vegetative cells of the control, i.e. parental strains in monoculture, progressed through the cell cycle, while both MTs in the glass bipartite apparatus arrested their cell cycle in the G1. Many

more strongly regulated genes were found in MT- than in MT+; in Scalco et al. (2016), the microscopic and time-lapse observations showed that the first cell in pairs that undergoes meiosis is the MT- suggesting that the response to mate perception is slightly delayed in MT+ and so partially explaining the different gene regulation between opposite MTs. Genes differentially expressed during early phases of sexual reproduction are mainly involved in signalling, metabolism, nutrient transport and meiosis. Of the 1,112 regulated genes, many of those that change in the earlier time point were regulated in both MTs, and in particular those annotated like nutrient transporters were downregulated. Two cyclins, homologues of *dsCYC5* and *dsCYC4* found in *P. tricornutum* (Huysman et al. 2010), were downregulated in the second time point. At the same time, genes encoding for the cohesin complex, required for the separation of sister chromatids during cell division (Patil et al. 2015), were upregulated in both MTs, indicating the preparation for meiosis. Many other strong regulations were detected, like a strong upregulation of gene encoding for a soluble guanylate cyclase, or upregulation of a gene involved in the inositol phospholipid signalling pathway. Among these genes differentially expressed, 11 resulted under positive selection. Of these, six were not annotated, two were peptidases, two leucine-rich repeat (LRR) receptor-like protein kinases and one putative DNA helicase (Basu et al. 2017).

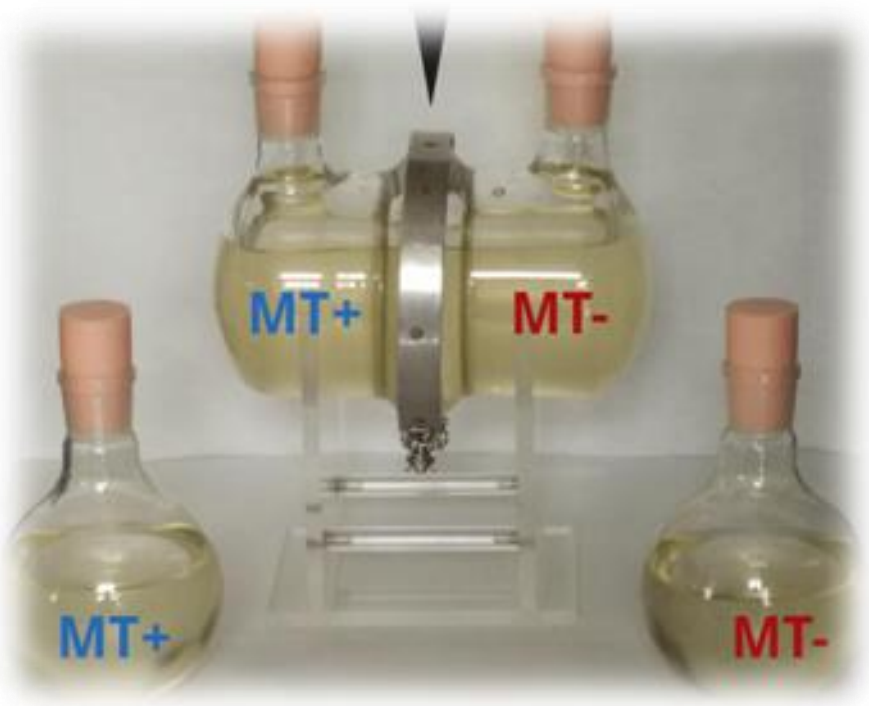


Fig.1.5 Bipartite glass apparatus used in Basu et al., (2017). This system allows to put in co-culture two opposite MTs for studing the chemical communication between separate MTs.

Another study was aimed at profiling genes differentially expressed between the two MTs, with the ultimate goal of identifying the gene responsible for MT determination (Russo et al. 2018). In this study, the transcriptome profiles of three MT+ and of three MT- were compared and 35 differentially expressed genes were found. Among these genes, 5 were of great interest for their putative implication in sexual reproduction and in sex determination. Three, *MRP1, MRP2* and *MRP3* (Mating type Related Plus) were more expressed in MT+, and two, *MRM1* and *MRM2* (Mating type Related Minus) were more expressed in MT-. Using prediction software and manual annotation, the function of the 5 protein products of the MR genes were predicted. MRP1 contains a putative signal peptide indicating that the protein might be involved in secretion or targeted to specific organelles. MRP2 and MRM2 were predicted to have leucine-rich repeat (LRR) receptor-like protein kinase domain and a transmembrane region. A more accurate manual inspection of the predicted protein sequence of these two genes did not confirm the presence of the kinase domain. LRRs are

20–29 amino acid motifs present in several proteins with diverse functions and cellular locations. These repeats are usually involved in protein - protein interactions; when acting serially they form nonglobular, crescent-shaped structures (Kobe and Kajava 2001). The LRR domains are present in brassinosteroid (steroid hormone) receptors and furthermore in animals the steroid hormone receptors are ligand-activated transcription factors (Torii 2004). MRM1 has a Heat Shock Factor (HSF)-type DNA-binding domain. Finally, MRP3 did not show any recognisable domain. The authors focalized their attention on *MRP3*, discovering that this gene shows several MT specific allelic variants in the upstream region: alleles A (Alto, high), M (Medio, medium), B (Basso, low) and N (Niente, nothing) (Figure 1.6).

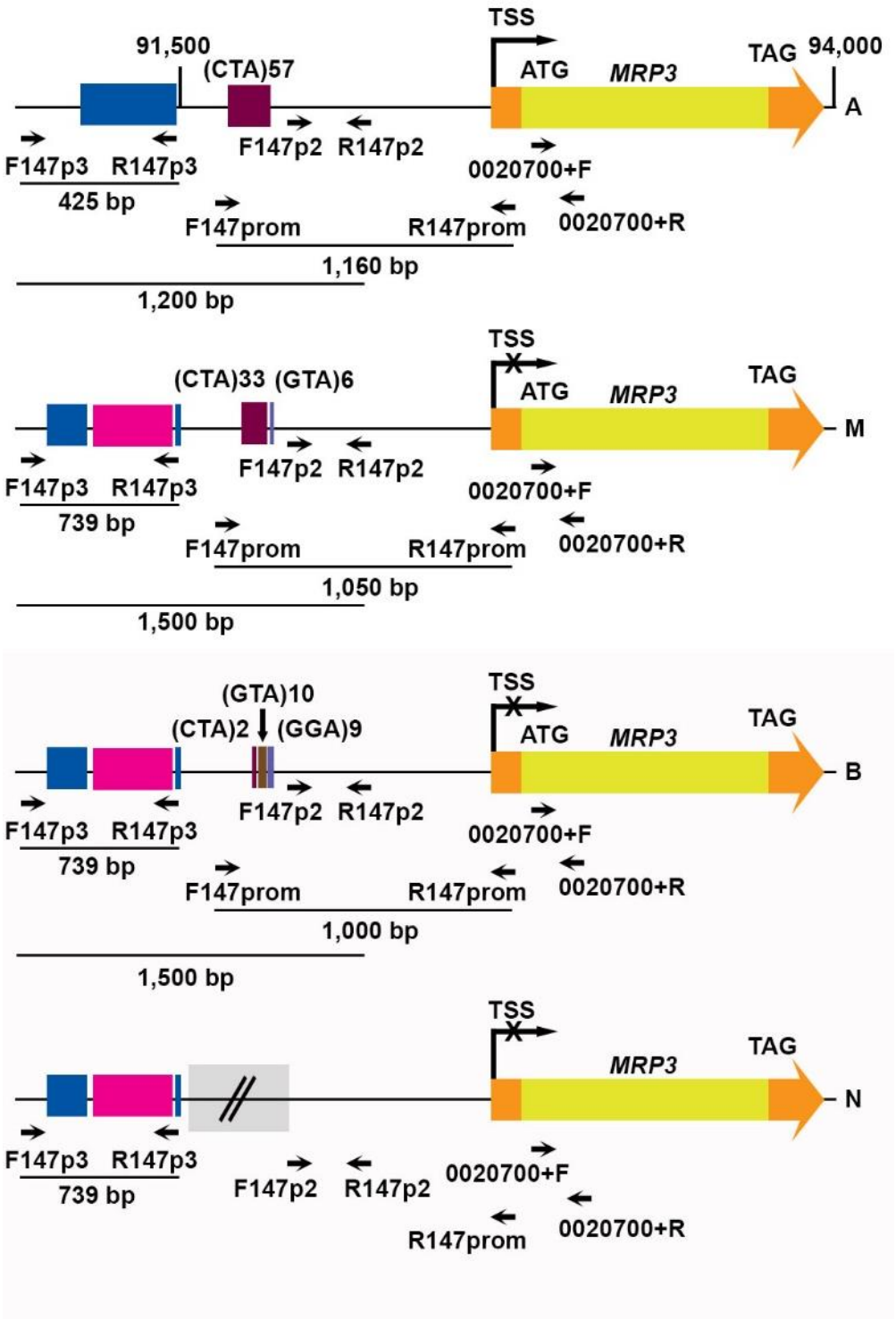


Fig.1.6 Schematic representation of the A, M, B and N alleles. The *MRP3* gene is represented like a yellow arrow with 5' and 3' UTR in orange. The purple, brown and violet boxes represent CTA repeats, GTA repeats and GGA repeats respectively, the number of repetitions is indicated for each triplet. The pink box represents a repetitive sequence, the blue box remnants of a transposase. The sequence in the gray box could not be amplified. Cartoon from Russo et al. (2018).

Among these allelic variants the allele A was present only in MT+ and never found in MT-. These alleles present a variable number of CTA repetitions, the presence of GTA triplets in the M and B alleles, and GGA triplets in the B allele in upstream region of *MRP3* gene (Fig. 1.6). Finally, overexpression of *MRP3* induced sex reversal in two different MT- strains; indeed, they were able to mate with another MT- and not with MT+. Furthermore, the expression of *MRP3* in MT- affected the other MR genes: the MRMs genes were repressed, while MRPs genes were upregulated respect to wild type MT-. These experiments demonstrated that *MRP3* is the mating type determining gene. All these results throw light on an essential step of the life cycle of *P. multistriata* that becomes an ideal reference candidate for studies in mate perception and sexual reproduction.

1.10 Aims of the thesis

The fascinating questions about the process of diatom sexual reproduction at sea are many, but to me the most important ones are to understand in which way the cell perceives the cues, how it responds to the signals, and finally in which way opposite MTs communicate between them to maximise the reproductive success.

The *Pseudo-nitzschia multistriata* life cycle is characterized by cell size reduction during vegetative growth and cell size restitution during the sexual phase. Experimental data suggest that the switch from vegetative division to sexual reproduction is mediated by chemical cues. This chemical cross talk could be a short (up to 6-8 hours) and speedy exchange of few molecular messages between MTs or could be a slow and long (up to 24-30 hours) cross talk that lasts until gametes production, or even longer. In both possible scenarios for this planktonic species there must be a first molecular signal from one of the two opposite MTs, and perhaps a molecular signal for attraction of the opposite MT. The available resources for this diatom, including a sequenced genome, transcriptomic data and genetic manipulation protocols, make *P. multistriata* an ideal model to explore the cross talk during mate perception and the mechanisms that regulate sexual reproduction. The aims of the thesis were:

1. Identify genes that respond directly to primary cues and genes that are activated only when the cross talk occurs. The first group of genes could be used as molecular markers to test fractions of the medium for a bioassay-guided identification of putative primary pheromones involved in the first part of the mate recognition system.
2. Elucidate the activation dynamics of a panel of selected genes during the early phase of sexual reproduction and possibly define the hierarchy of action.
3. Test the possible use of a panel of genes involved in the process to be used as markers for sexual reproduction in *in situ* data, exploiting the TARA Oceans global

expedition metatranscriptomics database. I tried to understand if these genes could be used to detect sex events in the sea.

4. Characterize metabolic differences between cells in the vegetative phase and cells in the sexual phase.

Chapter II illustrates mainly the results of gene expression studies performed on several experimental conditions, to address points 1 and 2; these conditions were:

- Each MT treated with the filtered medium of the opposite MT.
- The opposite MTs put in co-culture in a glass bipartite apparatus, allowing the chemical communication between them without physical contact, to perform a time course of target genes expression for both MTs.
- The opposite MTs put in co-culture in small scale, simulating the glass bipartite apparatus experiment, where cells were treated with chemical compounds to perturb mate perception.
- The opposite MTs co-cultured and followed for consecutive days (1 hour, 24 hours and 5 days)

Chapter III illustrates the results of searches in the TARA database for the five MT related genes and two meiosis genes. The results were used to perform phylogenetic analyses and a small sub set of genes in the trees were mapped on the TARA stations using geographical coordinates to display the distribution patterns.

Chapter IV illustrates the results of metabolomics analyses of the opposite MTs growing as monoculures and in co-culture, when sexual stages are present.

Chapter V presents a summary of the main conclusions and of future perspectives

Chapter II:

Cross talk between mating types: transcriptional changes and mating dynamics

2.1 Introduction

Recent evidence shows that in heterothallic pennate diatoms, sexual reproduction involves many signalling mechanisms between opposite MTs that may be quite different between species (Frenkel et al. 2014). However, there is also increasing evidence that the first steps of the sexual reproduction are similar in multiple species, and that as yet unidentified pheromones play a significant role for the synchronization of mating; resulting in the synchronised production of gametes thus ensuring a high reproductive success.

The species *Pseudostaurosira trainorii* produces motile male (MT+) and non-motile female (MT) gametes (anisogamy) (Sato et al. 2011). The MT+ gametes move randomly, thanks to structures called 'threads', but when they are close to MT- gametes they move towards the MT- in an amoeboid way. As mentioned before (see Chapter I the section 1.7), the mating is mediated by three putative pheromones: the first MT- sex pheromone induces sexualisation of MT+ that produces gametes and a second pheromone, which induces gametes production in MT-, that in turns produces a putative third pheromone that is an attractant for MT+ gametes. This signalling cascade occurs in a light-independent way. Indeed, sexual reproduction in this species takes place also in dark conditions.

In *Seminavis robusta*, gametes formation occurs only after pairing of opposite MTs that is strictly light dependent (Gillard et al. 2013). Both MT- and MT+ cells produce a sex-inducing pheromone (SIP-, SIP+, respectively); these compounds trigger cell cycle arrest in G1 phase in the other mating type and in MT- cells, the perception of SIP+ leads to the release of a third pheromone by MT-cells, diproline. that attracts MT+ (Figure 2.1).

Cell cycle arrest is associated with the downregulation of the expression of two mitosis-related genes that are orthologues of *CYCB1* and *CYCA/B1*, two cyclins highly expressed during mitosis in the diatom *Phaeodactylum tricornutum* (Huysman et al. 2010). These genes are expressed during mitosis in untreated *S. robusta* cells, while their expression is repressed when the MT- cells are treated with SIP+ (Moeys et al. 2016).

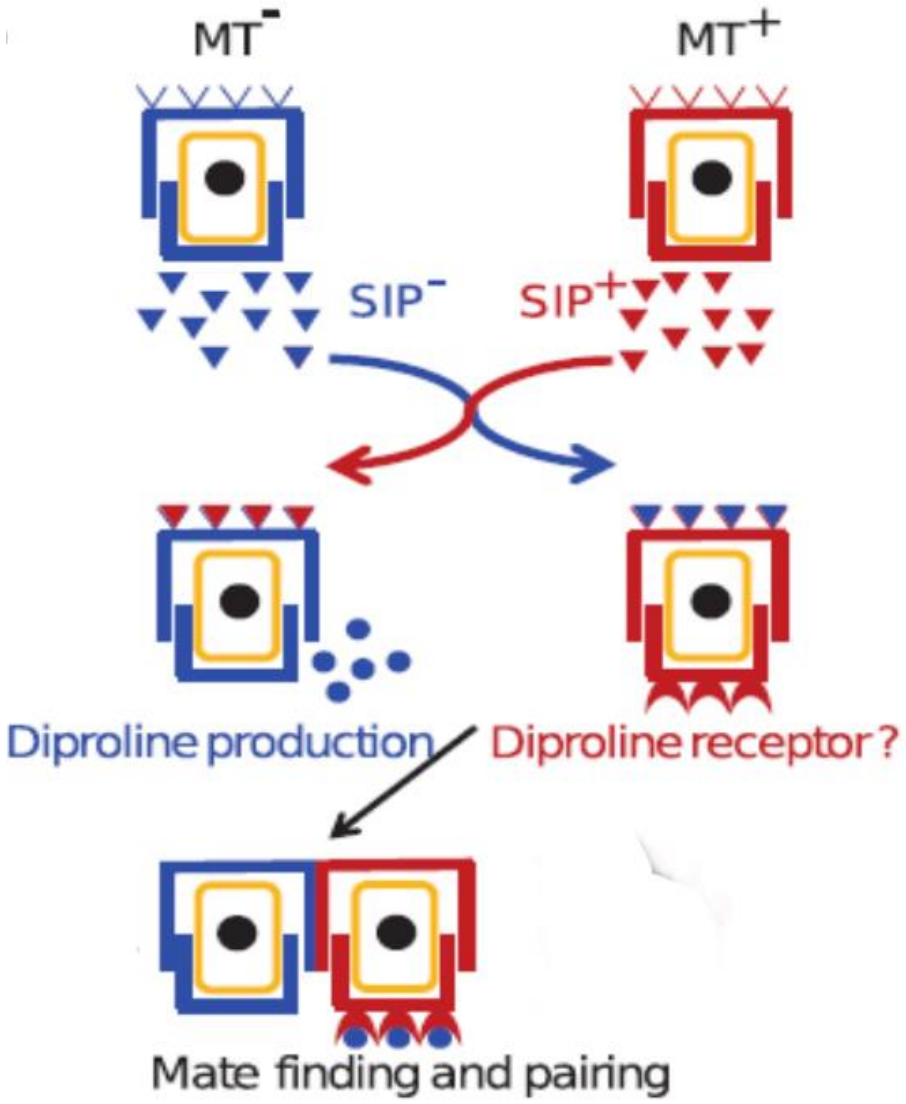


Fig.2.1 MT+ and MT- of *S. robusta* below SST during mating: both MTs produce sex-inducing pheromones, SIP+ and SIP−. SIP+ induces the secretion of the attraction pheromone diproline in MT− that is bound by a putative diproline receptor, allowing mating and pair formation. Figure from Moeys et al. (2016).

Results from RNAseq data of MT- cells showed that four meiosis-related genes were highly expressed in SIP+-treated cultures as compared to the untreated control (Moeys et al. 2016). Two genes coding for enzymes involved in diproline production were upregulated in treated cultures with a strong increase in expression over time, and finally, the expression of a gene, annotated as bi-functional guanylyl cyclase/phosphodiesterase, in treated cultures, increased over time (15 minutes, 1 hour and 3 hours). The guanylyl cyclase domain is responsible for cyclic nucleotides synthesis (cGMPs, second messengers involved in signal transduction) and the phosphodiesterase domain is involved in the breakdown of them. It is reasonable to think that this bi-functional protein has a key role in pheromone signalling in *S. robusta*. While mate recognition system has sofar been mainly explored in benthic species, it is uncertain whether it also applies to planktonic diatoms who live in a turbulent environment where chemical trails are easily dispersed.

A model species for investigating the mate recognition system in planktonic species is *Pseudo-nitzschia multistriata*. RNAseq data of the phase in which gametangia pair before meiosis (Basu et al. 2017) represent a good reference dataset to approach the study of the signalling processes occurring when two opposite MTs below SST perceive each other and to compare with SR data. The experiment was carried out by placing the MT+ and MT- cultures in two bottles separated by a filter that allowed chemical communication but not the contact between opposite mating types. From this study, it emerged also that, once the chemical signal of the other MT is perceived, the majority of cells arrest their cell cycle in G1 (Basu et al. 2017). Gene expression changes are reported mainly in the inositol pathway, in meiosis and in nitrate and ammonium transporters (Figure 2.2).

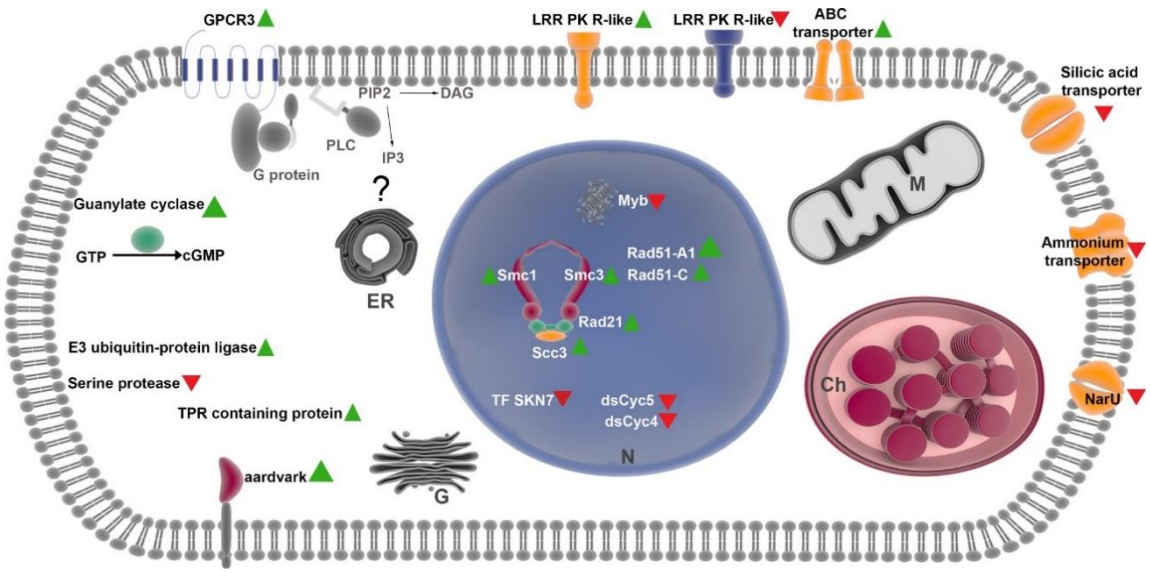


Fig.2.2. Cell response to sexual cues. Diagrammatic representation of a *Pseudo-nitzschia multistriata* cell with the principal genes involved in the response to chemical cues acting at the beginning of sexual reproduction. Green triangles represent upregulation and red triangles downregulation of expression. PLC, Phospholipase C; DAG, diacylglycerol; PIP2, phosphatidylinositol biphosphate; IP3, inositol trisphosphate; GTP, Guanosine-50-triphosphate; N, nucleus; ER, endoplasmic reticulum; M, mitochondrion; Ch, chloroplast; G, Golgi; LRR, leucine-rich repeat (Basu et al. 2017).

In addition, genes annotated as G protein-coupled receptors, kinases and a soluble guanylate cyclase were also upregulated when opposite MTs perceived each other. Soluble guanylate cyclase synthetizes cGMP. In metazoans, cGMP plays a key role in perception of pheromones and it is synthesized by both membrane and soluble guanylyl cyclase (Alberts 2015). The soluble guanylyl cyclase is activated by nitric oxide (NO) that binds to its iron group. NO is a second messenger that can act inside the cell but it can also act outside it, passing across the membrane and transporting the signal to neighbouring cells. In eukaryotic cells, NO is produced by the nitric oxide syntheses (NOS). In mammals, there are three kinds of NOS: neuronal NOS (nNOS), endothelial NOS (eNOS) and inducible NOS (iNOS); nNOS and eNOS are constitutively expressed while the expression of iNOS can be induced by bacterial lipopolysaccharide, cytokines, and other agents (Forstermann and Sessa 2012). The NOS activity can be stopped using 1-[2-(Trifluoromethyl) phenyl imidazole] (TRIM) (Handy et al. 1996). A NOS transcript was

found in *P. multistriata* (Di Dato et al. 2015). Other secondary messengers that could be involved in the sexual reproduction signaling are diacylglycerol (DAG) and inositol 1, 4, 5 triphosphate (IP3). IP3 is produced at the plasma membrane level, it diffuses in the cytosol and binds to InsP3 receptors located on the endoplasmic reticulum, triggering the opening of Ca2+ channels and the release of Ca2+ into the cytoplasm (Gilroy, Read, and Trewavas 1990; Dodd, Kudla, and Sanders 2010; Alberts 2015) . DAG and Ca2+ activate the protein kinase C (PKC). DAG can be cleaved to obtain arachidonic acid that is both a second messenger and the substrate to produce other small lipids. In *P. multistriata*, the above mentioned pathways could be involved in mate recognition, as emerged from the RNAseq study that showed transcriptional changes in the inositol phospholipid signalling pathway, which is downstream of GPCRs, and in cGMP synthesis (Basu et al. 2017).

A strategy to reconstruct the hierarchy of molecular events occurring in response to sexual pheromones can be to target several hypothetical steps using chemical compounds (drugs like inhibitors or analogues) to perturb these steps. Perturbing a specific signalling pathway will cause molecular changes that may be downstream of pheromone perception and thus alter sexual reproduction. In literature there are several studies where inhibitors were tested in marine organisms. For example, in Shikata et al. (2011) the light-induced germination of resting spores in the centric diatom *Leptocylindrus danicus* was investigated. It was shown that LY83583, that inhibits guanylyl cyclase, inhibited spore germination under one of the three photoperiods used in the experimental setup. LY83583 inhibits guanylyl cyclase blocking the production of cGMP with consequences on intracellular Ca2+.

In another study on *Sepia officinalis*, several inhibitors were used to understand the role of NO in the chromatophore organs response. In particular, the NO donor, that is 2-(N, N diethylamino) diazenolate 2 oxide (DEA/NO) induced a highly significant chromatophore expansion. Furthermore, NO-induced chromatophore expansion was arrested by treatment with 1 hour [1,2,4] oxadiazolo[4,3-a] quinoxalin-1one (ODQ) that is a soluble guanylyl

cyclase inhibitor, whereas using 8-bromo-cyclic GMP (8-Br-cGMP), that is a membrane permeable compound analogue of cGMP, with resistance to phosphodiesterase activity, there was a significant chromatophore expansion. This result showed the involvement of NO as an important messenger in body coloration in *Sepia* (Mattiello et al. 2010)

In this chapter, I describe the results of several experiments I performed to elucidate the molecular interaction between MT+ and MT-. The preliminary step was to select a set of genes that were up or downregulated during the early phase of sexual reproduction (Basu et al. 2017, Russo et al. 2018). I chose genes involved in several biological aspects:

- Sex determination and specification of the mating type
- Nutrient up-take
- Potential pheromone degradation, cell adhesion and signal transduction
- Cell cycle and meiosis

The expression profile of these genes was investigated to verify which of them were induced and supressed in the different experimental conditions. The experiments aimed to address the following questions:

i) Is there an MT that initiates the sexualisation, as in *P. trainorii*, or are there pheromones constitutively secreted by both MTs below the SST, as in *S. robusta*?

ii) What genes can be used as molecular readout in bioassays to isolate putative pheromones?

iii) Among selected genes, what are the major genes most involved during mate perception and what is the temporal activation of them? What is the relationship among genes?

To respond to first question (i) I performed experiments where each MT was treated with the filtered medium of the opposite MT; this first approach allowed me understand which of the MTs reacts to putative pheromones contained in the filtered medium of the opposite

MT, inducing the expression of target genes known to be involved in sexual reproduction (Basu et al. 2017). Furthermore, this approach allowed me to identify a gene that can be used as molecular proxy to isolate the putative pheromones(ii).

The second approach, a time course of gene expression in sexualized MTs, allowed me to define the steps of cross talks in terms of expression changes of selected genes, identifying what are genes mainly involved in the mate perception. (iii); while to understand the relationship among genes (iii) a third kind of experiments, called “perturbation experiments”, were performed. Perturbation experiments were performed using specific inhibitors to interfere with specific putative pathways involved in the regulation of selected genes, to understand the hierarchy of expression of investigated genes.

The last section reports experiments aimed to understand if one of the two MTs can affect the mate efficacy, using crosses with different MT ratios, giving additional indications on the dynamics of mate perception.

Finally, I analyse the expression changes of target genes in RNAseq data recently made available in the laboratory. This RNAseq dataset was produced considering a wide time window of the sexual phase where MT + and MT- were in co-culture, up to 5 days, and it allowed to draw a more complete picture of the gene profiles.

2.2 Materials and Methods

2.2.1 Cultures

In Table 2.1 are indicated the pairs of strains used in the different experiments. The strain MC1217-17 was isolated at the LTER-MC station in Gulf of Naples in 2017 while all the other strains were obtained by crosses carried out in the laboratory. The cultures were grown in f/2 culture medium (Guillard, 1962) prepared with oligotrophic seawater. Strains were maintained in a growth chamber at a temperature of 18 °C, a photoperiod of 12:12 hours Light: Dark and an irradiance of 90 µmol photon m^{-2} s^{-1}.

Tab.2.1. Pairs of strains used in the different experiements, for each strain the cell size at the time in which experiments were carried out is indicated.

	Couples	
Experiments	**MT+**	**MT-**
Conditioned medium experiments	LV168	LV92A5
	24 µm	24 µm
Time course of sexualized MTs	LV130	LV193
	30 µm	30 µm
Perturbation experiments	Pmf3.2	LV92A5
	25 µm	26 µm
Sex ratio experiments	LV80	MC1217 (17)
	30 µm	18 µm
	LV168	CB-B3
	21 µm	45 µm
	LV130	CB-B3
	30 µm	45 µm

2.2.2 Cell synchronization

For all the experiments the cell cycle of strains was synchronized by keeping cells in the dark for 36 hours. The cell synchronization started when the cell culture was in the exponential growth, with cell density ranging between 1.0-1.5 × 10^5 cells mL^{-1}.

2.2.3 RNA extraction and reverse transcription

Cells were collected by filtration on 1.2 µm pore size nitrocellulose membranes RAWP04700 or RAWP01300 (Millipore). Filters were submerged in a variable volume of TRIzol® (1 ml for 5-10 x 10^6 cells), flash frozen in liquid nitrogen and stored at -80 °C. RNA extraction was performed using TRIzol® (Roche, Basel, Switzerland). After defrosting, acid washed glass beads (Sigma) were added to samples, which were incubated in a thermo-shaker at 60 ^{0}C for 10 minutes at maximum speed. After recovering the supernatant, chloroform was added (1/5 of the volume) to each sample, mixed well by shaking vigorously for 15 seconds and incubated for 15 minutes at room temperature to separate the phases. For a clearer separation of phases, the samples were centrifuged for 15 minutes at 12,000 g at 4 ^{0}C. The uppermost aqueous layer (containing RNA) was transferred into a new tube without disturbing the middle colourless phase and the lowermost pink phase (these phases contain proteins and DNA). At the aqueous phase was added an equal volume of isopropanol; the samples were gently mixed, incubated at room temperature for 10 minutes, and then centrifuged for 10 minutes at 12,000 g at 4 ^{0}C. The supernatant was removed leaving the RNA pellet undisturbed. The pellet was washed with 1 ml of 75% ethanol mixed the tube gently, and then centrifuged for 10 minutes at 8,000 g at 4 ^{0}C. The supernatant was removed leaving the RNA pellet undisturbed. The pellet was dried for about 20 minutes to remove all the traces of ethanol and then re-suspended in 12 or 30 µL of DEPC water. Genomic DNA contamination was eliminated digesting with DNase I (Qiagen) according to the manufacturer's instructions, followed by RNA purification using sodium acetate according to Molecular Cloning (Wood 1983). The RNA

was analysed by gel electrophoresis (2% agarose w/v), the Qubit® 2.0 Fluorometer (Life Technologies) was used to assess concentration and the NANODROP (ND 1000) spectrophotometer to determine the quality as 260/280 nm and 260/230 nm absorbance ratios. RNA contamination by genomic DNA was tested with PCR amplification, using the RNA as template. A subsample of 200-500 ng of the total RNA extracted was used for cDNA preparation using the QuantiTect® Reverse Transcription Kit (Qiagen). cDNA quality was assessed amplifying a 177 bp fragment containing an intron of the cyclin dependent kinase (CDK A) gene with primers *CDK Pm fw* 5′-GTGCACACGGAACGGAAACTC-3′ and *CDK Pm rv* 5′-CTTCAAATCTCTGTGCAGTAC-3′.

2.2.4 qPCR and data analysis

Real time qPCR amplification was performed using 1 µL of a diluted cDNA, 4 µL of the primers (final concentration 0.7 µM of each primer) and 5 µL of Fast SYBR Green Master mix with ROX (Applied Biosystems) in a final volume of 10 µL, using ViiA™ 7 Real-Time PCR System (Applied Biosystems). Each sample was analyzed in technical triplicate to capture intra-assay variability and each assay included at least two negative controls for each primer pair. PCR conditions were as follows: 95 °C for 20 s, 40 cycles at 95 °C for 1 s and 60 °C for 20 s, 95 °C for 15 s, 60 °C 1 minute, and a gradient from 60 °C to 95 °C for 15 minutes. To measure the relative expression, the method of -ΔCT was used (Lohbeck et al. 2014). The relative expression of a specific gene for each sample was calculated using normalized values with respect to the geometric mean of reference genes. The -ΔCT was equal to the difference in threshold cycles (CT) for reference and target genes (geometric mean of reference genes CT -target CT). The results were transformed in fold expression change (FC) equal to $\pm 2^{(|(-\Delta CT\_Treatment) - (-\Delta CT\_Control)|)}$. The reference genes used for qPCR were *TUB A*, *TUB B* and *COP A* (Adelfi et al. 2014), and the best

combination of reference genes was chosen using the NormFinder software (Andersen et al. 2004). Primers for the genes tested in the study are listed in Table 2.2.

Tab.2.2. Primers for the *Pseudo-nitzschia multistriata* genes used in this study.

Gene name	Gene ID	Primer name	Sequence 5'→3'	Amplicon size (bp)
NarU	PSNMU-V1.4_AUG-EV-PASAV3_0048930.1	PM_0048930 F	CACGTCGCTGCTACTCTGTA	178
		PM_0048930 R	TGCCATTCCTTGAAGACGCT	
75120	PSNMU-V1.4_AUG-EV-PASAV3_0075120.1	PM_0075120 F	AGGAGCCAATGAAGGTGACG	179
		PM_0075120 R	GCCTTCCCTGACTGTTGTGA	
sGC	PSNMU-V1.4_AUG-EV-PASAV3_0051110.1	PM_0051110 F	GAACGTGGTGCCAGCTTTTG	198
		PM_0051110 R	GTTGAGAAGGAAGCGGTCGA	
7488	PSNMU-V1.4_AUG-EV-PASAV3_0020420.1	comp7488_c0_seq1.1_F	AGCAAAGCCGACGATGCC	170
		comp7488_c0_seq1.1_R	AATTCGTGCGATTCTCCGTTG	
MRP1	PSNMU-V1.4_AUG-EV-PASAV3_0024820.1	0.00+ F	GTATGGCGCTCACCACTTC	156
		0.00+ R	CGTCTTCGACTGCGTCTTC	
MRP2	PSNMU-V1.4_AUG-EV-PASAV3_0122240.1	127.15 F3	CCTCCGAATATGGATACATG	194
		127.15 R3	GAGCTAAACATCGTGACACC	
MRM1	PSNMU-V1.4_AUG-EV-PASAV3_0085380.1	47507 F	CCCCTACAAGCTCTTTGATTTG	160
		47507 R	GAAATTGTGGTGCCCAAAG	
MRM2	PSNMU-V1.4_AUG-EV-PASAV3_0006960.1	46228 F	CCACCGAACTAGGCAACTGTC	139
		46228 R	GGCACAGAACCCGTCAAC	
MRE11	PSNMU-V1.4_AUG-EV-PASAV3_0086370.1	PM_0086370 F	TTGCGGTGTTTGATGTTCGT	214
		PM_0086370 R	TCCAAGAGCTCCCTTCGTTT	
RAD50	PSNMU-V1.4_AUG-EV-PASAV3_0001820.1	PM_001820 F	CGAAAGAAGCGCTACGACAA	246
		PM_001820 R	AACAACCCGGAAGTACTCGT	
CYCB1	PSNMU-V1.4_AUG-EV-PASAV3_0122920.1	PM_0122920 F	CAAGCGAATCGTGCAGCTAT	184
		PM_0122920 R	TGCAATCTCTTCCTCACGGT	
GCPDE	PSNMU-V1.4_AUG-EV-PASAV3_0076150.1	PM_0076150 F	CGTCGTGGTGTTTTCAAGGT	217
		PM_0076150 R	CAGTGACAGGTCCAGAGTGT	
Cathepsin D	PSNMU-V1.4_AUG-EV-PASAV3_0103000.1	XLOC_008922_F	CCATTGTCGACTCGGGAACC	170
		XLOC_008922_R	GTAGGTTTTCCCGCCGATCC	
Aardvark	PSNMU-V1.4_AUG-EV-PASAV3_0102760.1	PM_0102760_F	ACGACATCCACAGCAATCTC	149
		PM_0102760_R	TTCCTTGAGCCGTTCCAGAC	
31872	PSNMU-V1.4_AUG-EV-PASAV3_0001670.1	comp31872_c0_seq1.1_F	CGAGTTTCTCACATTTGCTGG	169
		comp31872_c0_seq1.1_R	CTCTTTTGGATGTGGACCG	
RAD51-A1	PSNMU-V1.4_AUG-EV-PASAV3_0056780.1	RAD51-A1 Pm F	CATCGGCGGGAACATCATTG	160
		RAD 51-A1 Pm R	ATCGGTAGCATCGCAAACTC	
CYCB2/2	PSNMU-V1.4_AUG-EV-PASAV3_0095090.1	Cycb1_F	ACAACAGGTCTTGAGAGCAC	163
		Cycb1_R	CGTGGTCTTTCGATTCCTTC	

2.2.5 Conditioned medium experiments

Three different experiments were carried out using the same pair of strains (Table 2.1) grown at the same growth condition and same final density (1.0-1.5 × 10^5 cells mL^{-1}). The conditioned medium experiments were conducted exposing MT+ and MT- strains for 2 and 6 hours to the filtered medium of the opposite MT. From now on, the cells exposed to the conditioned medium will be called "conditioned cells". The experiments included the controls, which were the two vegetative MTs in monoculture, and the MTs in the bipartite glass apparatus, i.e., separated by a filter, with the same set up described in Basu et al. (2017). Cells grown in the apparatus will be called "sexualized cells". The cells to be treated with medium conditioned by the opposite MT were filtered on a filtering unit of 500 ml (PES membrane 0.22 µm, TPP or Millipore) to concentrate them and were then re-suspended (Figure 2.3). The filtered medium was obtained from non-synchronized cells in exponential phase.

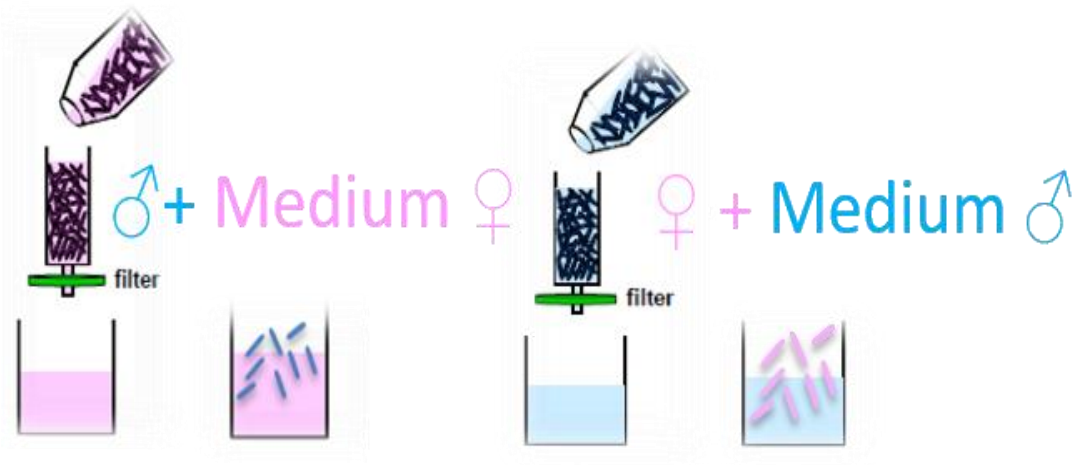


Fig. 2.3 Treatment with conditioned medium: cells of one MT were resuspended in the medium conditioned by the opposite MT.

In order to be sure that the filtration procedure did not affect the physiological state of the cells, the growth curve of cells concentrated and re-suspended in fresh medium was compared to the growth curve of a culture of non-concentrated cells. The two growth curves were comparable.

Samples for RNA isolation were collected from 200 ml of culture after 2 and 6 hours from the inoculum. A volume of 20 ml of treated cells was transferred in petri dishes to visually inspect the culture on a daily basis to test for the presence of sexual stages. In all experiments a cross test was set up to check if sexual reproduction occurred; the samples were processed for RNA isolation only in case of a positive response.

2.2.6 Time course of gene expression in sexualized MTs

Experiments were run using the same couple of strains (Table 2.1), the same growth condition and same cell density (1.0-1.5 × 10^5 cells mL^{-1}). Strains were grown in the bipartite glass apparatus and the individual strains were also grown in mono-culture at the same initial cell concentration. The two strains were synchronized as illustrated above and the chemical contact between MT+ and MT- started in the dark, and after 1 hour both controls and sexualized cells were exposed to light. The cells were collected at the following time points: 0, 30', 1 hour, 2, 3, 4, 5, 6 and 8 hours. RNA of samples at 2, 4, 6 and 8 hours were extracted as illustrated above and the expression of target genes - *MRM1*, *MRM2*, *MRP1*, *MRP2*, *7488*, *NarU* and *CYCB1* - was tested by qPCR.

2.2.7 Perturbation experiments with inhibitors

The compounds tested were: 2-(4-Carboxyphenyl)-4,4,5,5-tetramethylimidazoline-1-oxyl-3-oxide (CPTIO) (Alexis Biochemicals), 1-[2-(Trifluoromethyl) phenyl] imidazole (TRIM) (Sigma) and 6-Anilinoquinoline-5,8-quinone (LY-83583) (Sigma). The solvent for the first two compounds was dimethyl sulfoxide (DMSO), while for LY-83583 it was

methanol. Mock controls were made administering to the cells the same volume of solvent.

Toxicity tests were performed to identify the highest non-lethal concentration of the compounds selected for perturbation experiments. The tested concentrations for CPTIO were 10, 5 and 1 µM; for TRIM were 1, 0.1 and 0.01 mM; for LY-83583 were 2, 1, 0.2, 0.1, 0.01, 0.001 and 0.0005 µM. The highest non-lethal concentration of chemical compounds was used in a cross between the two MTs (Table 2.1) to check if sexual reproduction was affected by the chemical compounds. The experiment was designed to keep MT+ and MT- cells of *Pseudo-nitzschia multistriata* in chemical contact without physical contact and to perturb their chemical interaction during sexual reproduction. The perturbation experiments with drugs were set up in 6 multi-well culture plates (ThinCert™) in which each well was separated in two sub-compartments by a 1.0 µm pore sized membrane. The total volume of each compartment was 6 ml, distributed in 3 ml for MT + and 3 ml for MT -. For each inhibitor, 4 plates, each with 6 well, were set up: two plates with couples (cross with separating membrane) treated with the inhibitor, one with only MT+ and one with only MT-. The same set up with 6 plates was made for controls with solvent. Two cross tests were set up in small petri dishes to check if sexual reproduction occurred in presence of the solvent, and if sexual reproduction was blocked or delayed in the presence of the inhibitors; only if these two conditions were verified the samples were processed for RNA extraction. Samples were collected after 6 hours from the start of the experiment. The set up used for the treatments with inhibitors is shown in figure 2.4. For uniformity and for simplification, samples were collected after 6 hours only from the bottom of the multiwell.

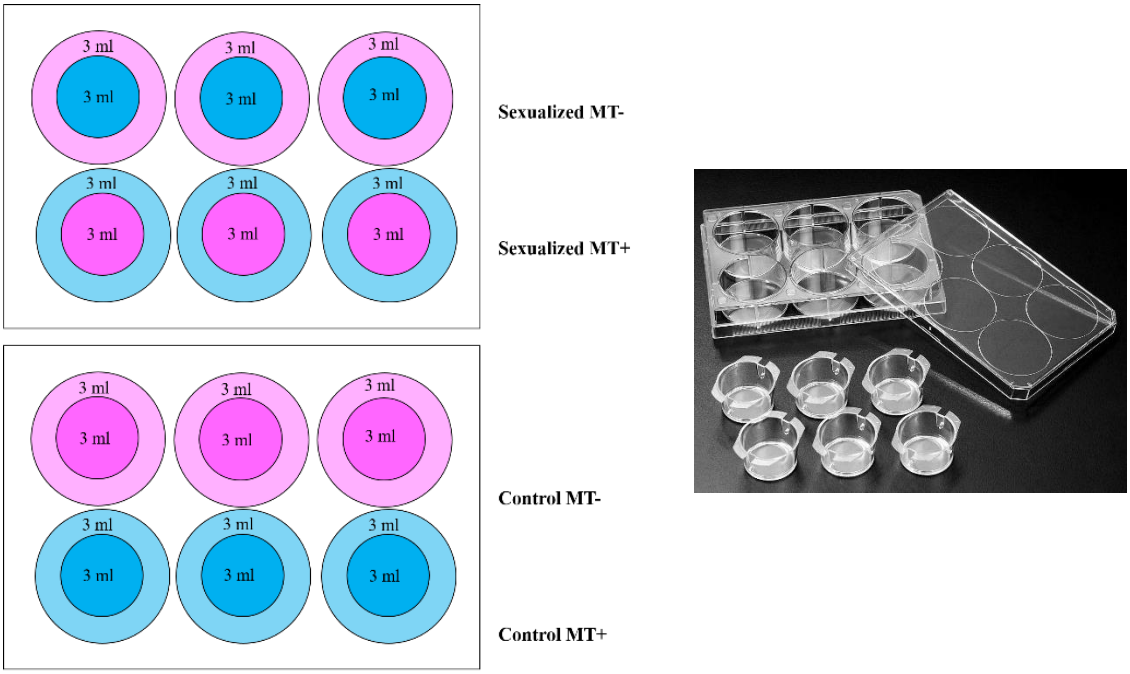


Fig.2.4 This cartoon shows the set up for treatements with inhibitors. a. The internal circle represents the insert menbrane placed in the well. Shades of pink and blu indicate respectively MT- and MT+ cultures. b. Example of culture plate with wells with insert membranes.

2.2.8 Sex ratio experiments

The crosses were set up between strains of different average cells size in a 6-well culture plate (Table 2.1). For the setup of the experiment the cells were collected when they were in a range concentration of 50-100 × 10^3 cells mL^-.

For each couple of tested strains, three different ratios of MT- and MT+ (as cell concentration) were tested, with three replicates each:

MT-/MT+: 50/50

MT-/MT+: 10/90

MT-/MT+: 90/10

The final total number of cells was 100,000 cells in 3 ml of medium. The crosses were observed using the inverted microscope equipped with a gridded ocular after 24 hours and mating events (pairs and clusters) were counted. The couples were: LV130(+) x CB-B3(-) (couple B), LV80(+) x MC1217-17(-) (couple C) and LV168(+) x CB-B3(-) (couple D). For each couple the experiment was performed two times.

2.2.9 Target genes expression profiles during later stages of the sexual reproduction process

In the laboratory, an experiment was designed and performed by Rossella Annunziata to characterize the gene expression of cells during a cross at three different time points: 1) cross after 1 hour of co-culture; 2) cross after 24 hours, in which some parental cells were in pairs and there was about 10% of gametes; 3) cross after 5 days with a heterogeneous population composed of parental MTs, few gametes, auxospores and about 2.5% of initial cells. Vegetative MT+ and MT- controls were set up in monoculture; final cell density at the beginning of the experiment was 20×10^3 cells mL^{-}. The experiment was performed using three replicates, always with the same couple of strains, LV130 (MT+) and LV193 (MT-). I exploited this dataset to follow the expression profile of my target genes.

2.3 Results

2.3.1 Choice of the target genes

To study mate perception of opposite MTs, I selected the most interesting genes from Basu et al. (2017) (Table 2.3). Mating type Related genes (MR genes) were selected since they may be involved directly in the pheromone-receptor system (Russo et al., 2018). *NarU* (nutrient up-take), annotated as nitrate/nitrite transporter, was chosen because it is downregulated in both MTs in the first phase of MT perception (6 hours), indicating a reduction of nutrient uptake. Cathepsin D, aardvark, *sGC*, *GCPDE* and 75120 were chosen for their possible implication in signalling. Cathepsin D could be involved in the degradation of polypeptide hormones; the aardvark protein could be related to cell differentiation for adhesion; *sGC* and 75120 were specifically chosen for their putative involvement in signal perception of odorant molecules. *GCPDE* was chosen for two reasons: it was highly upregulated in *S. robusta* MT- treated with SIP+, and it resulted under positive selection in *P. multistriata* (Basu et al. 2017) indicating that it could be implicated in reproductive isolation. The two unknown genes (7488 and 31872) were chosen for their strong upregulation during mate perception; in particular, 7488 seemed to increase only in MT- while 31872 in both MTs. *RAD51-A1*, *RAD50* and *MRE 11* were chosen because they are meiosis related genes (Patil et al. 2015), while *CYCB2/2* and *CYCB1* are cyclins, known to regulate the cell cycle (Huysman et al. 2010). In preliminary experiments (not shown) qPCR results for some of these genes (Cathepsin D, aardvark, *RAD51-A1, CYCB2/2*) indicated possible issues with the primer pair chosen or showed weak signals therefore they were excluded from some of the analyses presented in the following paragraphs.

Tab.2.3 Target genes chosen to study the process of mate perception in *P. multistriata*. The log FCs of both MTs are reported at two time points: 2 and 6 hours (Basu et al. 2017). In bold FCs below and above -1 /+ 1. For details on genes see the references; the genes are grouped based on their putative involvement in specific pathways or mechanisms: in peach genes of sex determination and specification of the mating type, in grey nutrient uptake, in pink genes involved in potential pheromone degradation, cell adhesion and signal transduction, in orange unknown genes, in light blue genes related cell cycle.

			T 2h		T 6h			
Gene Name	**Gene ID**	**Gene Function**	**MT+ S**	**MT- S**	**MT+ S**	**MT- S**	**Reference**	**Rationale**
MRP1	*24820*	Unknown	-0,36	-0,03	**3,60**	-0,84	Basu et al, 2017; Russo et al, 2018	These genes were chosen for their implication in the sexual determination process, moreover MRPs were specifically induced in MT+ while MRMs were specifically induced in MT- during mate perception
MRP2	*122240*	Leucine-rich repeat (LRR) trasmembrane region	0,75	0,19	**1,74**	**1,31**	Basu et al, 2017; Russo et al, 2018	
MRM1	*85380*	Heat shock factor protein	-0,17	**1,57**	-0,43	**3,40**	Basu et al, 2017; Russo et al, 2018	
MRM2	*6960*	Leucine-rich repeat (LRR) trasmembrane region	0,31	**1,85**	0,07	**2,55**	Basu et al, 2017; Russo et al, 2018	
NarU	*48930*	Nitrate/nitrite transporter NarU	**1,86**	0,98	**-2,42**	**-1,89**	Basu et al, 2017	This gene was downregulated in both MTs during sexual reproduction and its function could be related to metabolim regulation by modulating the uptake of nitrogen
Cathepsin D	*103000*	Cathepsin D	-0,71	**5,84**	0,42	**6,96**	Basu et al, 2017	These genes were chosen because they are part of the putative singling that occurred during sexual reproduction. Aardvark and Cathepsin D are putatively involved in pheromone degradation and cell-cell adhesion. sGC, GCPDE and 75120 are involved in general response to mate perception.
Aardvark	*102760*	Protein aardvark	**-1,68**	**5,37**	**6,09**	**3,81**	Basu et al, 2017	
sGC	*51110*	Soluble guanylate cyclase 88E	**1,35**	**5,77**	**6,30**	**4,99**	Basu et al, 2017	
GCPDE	*76150*	Guanylate cyclase with phosphodiesterase domain	-0,66	-0,16	**1,18**	**2,62**	Moeys et al, 2016	
75120	*75120*	Type II inositol 1 4 5-trisphosphate 5-phosphatase	0,08	0,04	**1,01**	**1,56**	Basu et al, 2017	
7488	*20420*	Unknown	0,16	**7,80**	-0,25	**6,80**	Basu et al, 2017	These unknown genes were strongly regulated during sexual reproduction, could be specie specific signals of this species
31872	*1670*	Unknown	0,39	**3,03**	**5,37**	**4,48**	Basu et al, 2017	
RAD51-A1	*56780*	DNA repair protein RAD51 homolog 1	0,62	**1,87**	**5,79**	**4,89**	Basu et al, 2017	RAD51-A1, RA50 and MRE11 were chosen for their role in meiosis, while CYCB2/2 and CYCB1 were chosen for their implication in regulation of cell cycle
RAD50	*1820*	DNA repair protein	0,85	-0,46	0,85	-0,46	Moeys et al, 2016	
MRE11	*86370*	Double-strand break repair protein	0,35	0,05	**1,31**	0,85	Moeys et al, 2016	
CYCB2/2	*95090*	Cyclin-B2-2	0,04	-0,46	**-2,34**	**-2,19**	Basu et al, 2017	
CYCB1	*122920*	Cyclin-B1	0,31	-0,54	**-1,91**	**-1,53**	Moeys et al, 2016	

2.3.2 Conditioned medium treatments

As detailed in the Materials and Methods, the filtered medium treatments were performed exposing synchronized MT cells to the filtered medium obtained from the opposite MT. The experiments included also the set up with bipartite glass chamber to have a picture of gene expression changes during sexualisation. This additional set up has two aims: i) compare the gene expression patterns obtained with a different pair of strains with those illustrated in Basu et al. (2017), and ii) have a positive control for the experiments carried out with the conditioned medium. In the first experimental set up I expect to see only the genes that respond to the perception of the opposite MT, while in the set up with the bi-partite apparatus I expected to have a more complex response that includes downstream steps after MT perception. These treatments were carried out to understand which selected genes are induced after the perception of putative pheromones contained in filtered medium, and if there is a gene (or more than one) that can be used as proxy for pheromone perception. The MT treated with the filtered medium from the opposite MT is called conditioned MT (MT +/- Cond). The same strains in the bipartite glass chamber apparatus are called sexualized MTs (MT +/– S). For each condition the results of three replicate experiments (009, 011 and 012) after two and six hours of treatment are shown in Figs 2.5-2.8, and the summary of the results are shown in tables 2.4 and 2.5. In table 2.4 for reference FC values obtained for the different genes in the experiments illustrated in Basu et al. (2017) are shown.

The response was different among replicate and time points. In the experiment in which the MT+ strains experienced the chemical signal through the membrane between the bottles *MRP1* was strongly induced at 6 hours, 7488 was strongly downregulated (2 hours) and then upregulated (6 hours), 75120 and *GCPDE* were upregulated at 6 hours, and finally *MRE11* and *RAD 50* were upregulated at 6 hours (Fig. 2.5).

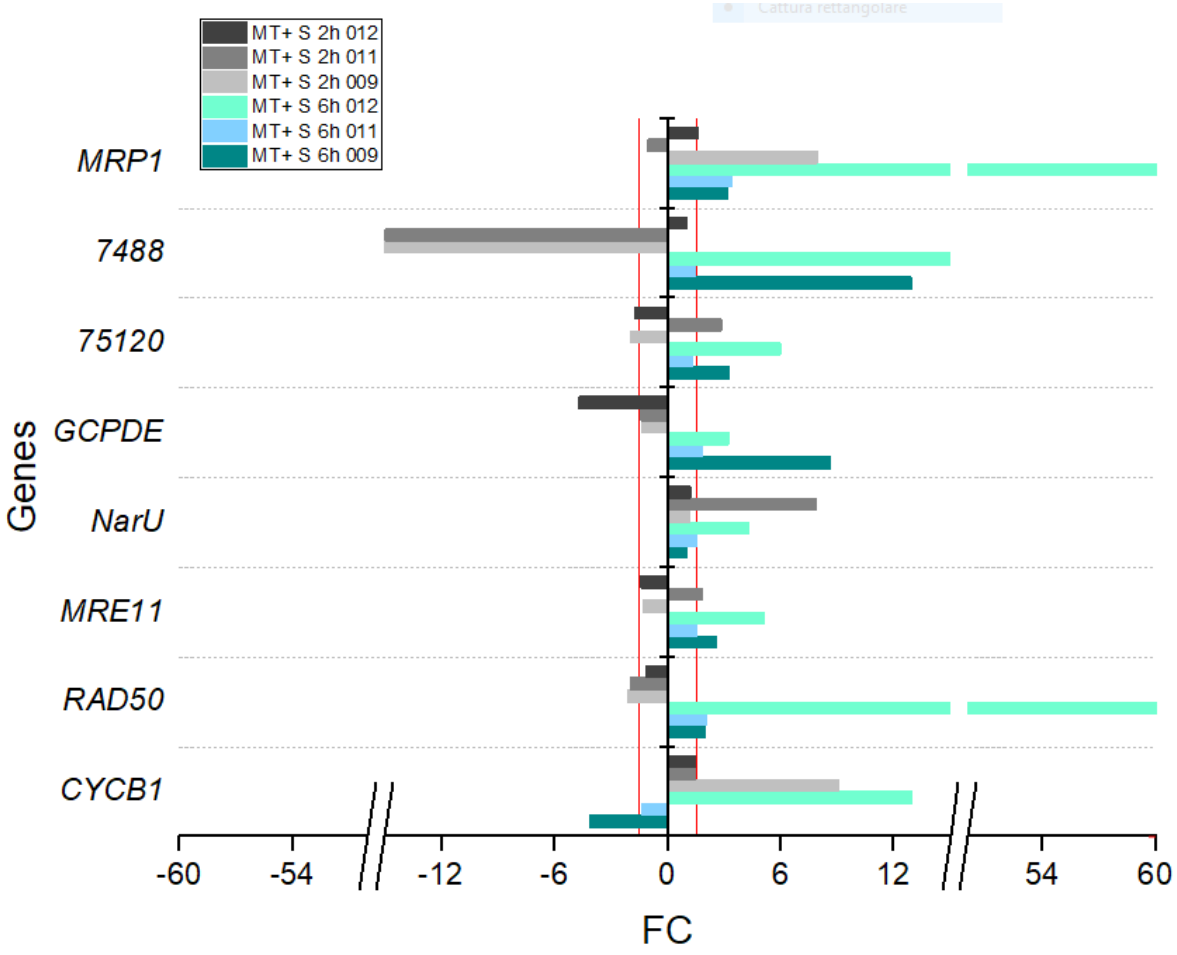


Fig.2.5 Fold change (FC) of the target genes in MT+ sexualized samples (MT+ S) with respect to the control in mono-culture. FCs are shown for the MT+ S at 2h and 6h, for all three experiments. The red line represents the significance threshold of FC equal to -1,5 and + 1,5.

In the 'Conditioned medium' set-up, in MT+ Cond the only gene strongly induced was 7488 at 6 hours (Fig. 2.6).

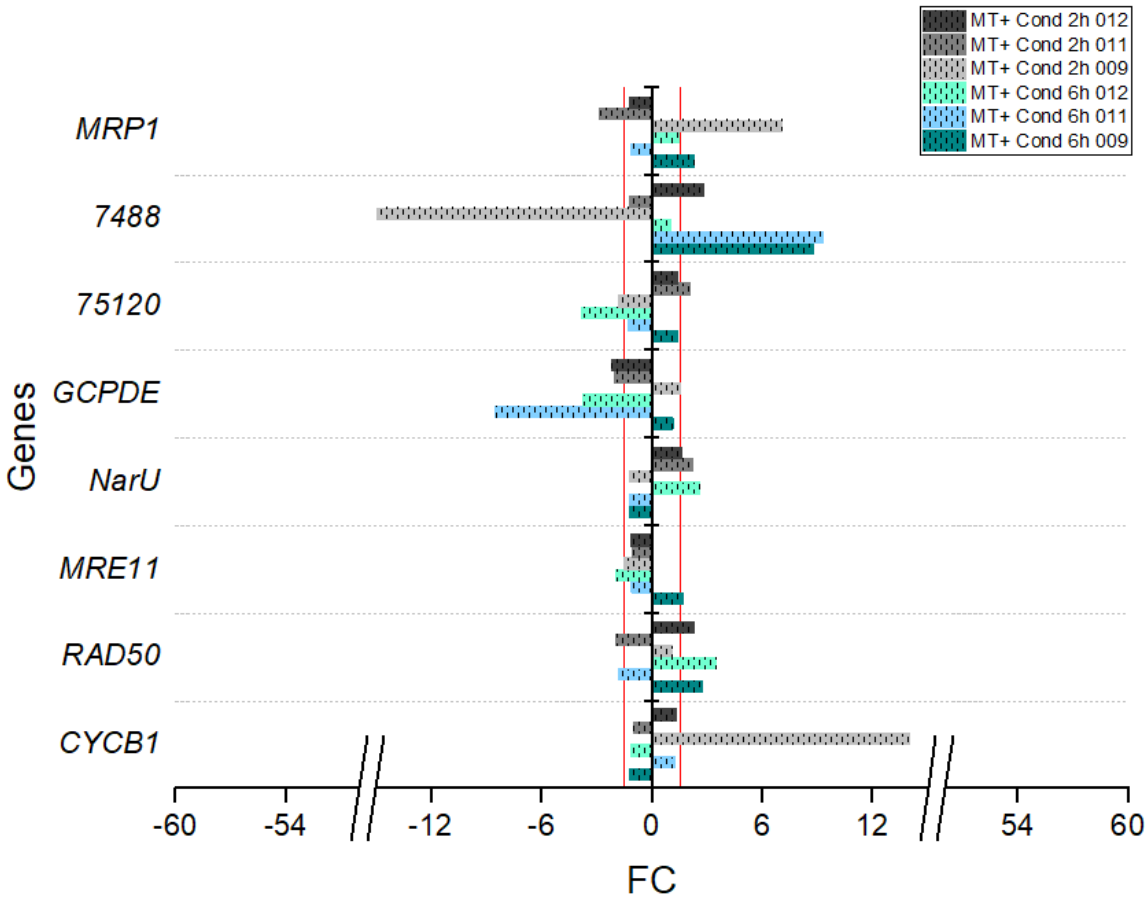


Fig.2.6 Fold change (FC) of the target genes in MT+ strains grown in conditioned medium (MT+ Cond) with respect to the control in mono-culture. FCs are shown for the MT+ Cond at 2h and 6h, for all three experiments. The red line represents the threshold of FC equal to -1,5 and + 1,5.

In MT- S the genes regulated were 75120, *GCPDE*, *NarU*, *MRE11* and *RAD50* at 2 hours (Fig.2.7); in MT- Cond, at both time points, *7488* was strongly upregulated, *CYC1B* was strongly downregulated and *MRE11* was upregulated. 75120 and *RAD50* were upregulated at 2 hours, while *GCPDE* at 6 hours (Fig.2.8).

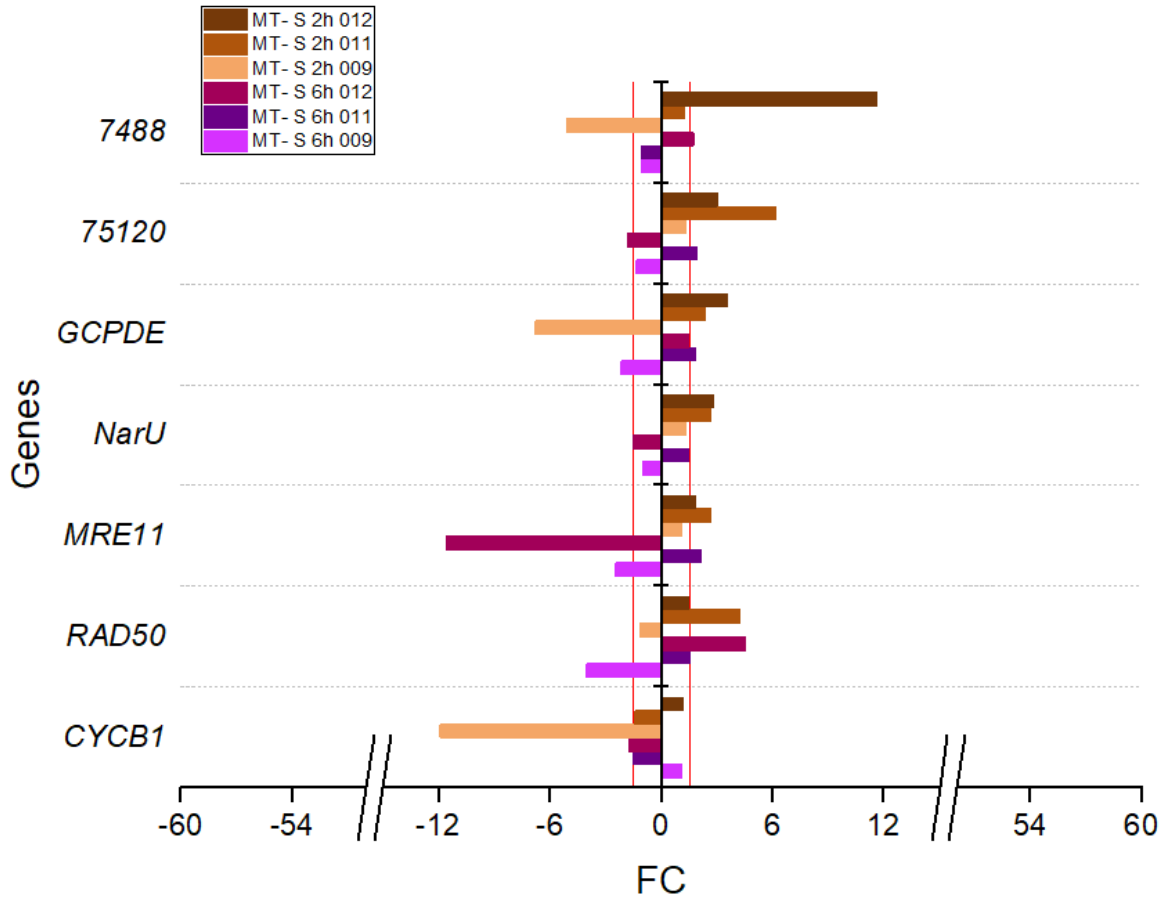


Fig.2.7 Fold change (FC) of the target genes in MT- Sexualized samples with respect to the vegetative control FCs are shown for the MT+ S at 2h and 6h, for all three experiments. The red line represents the threshold of FC equal to -1,5 and + 1,5.

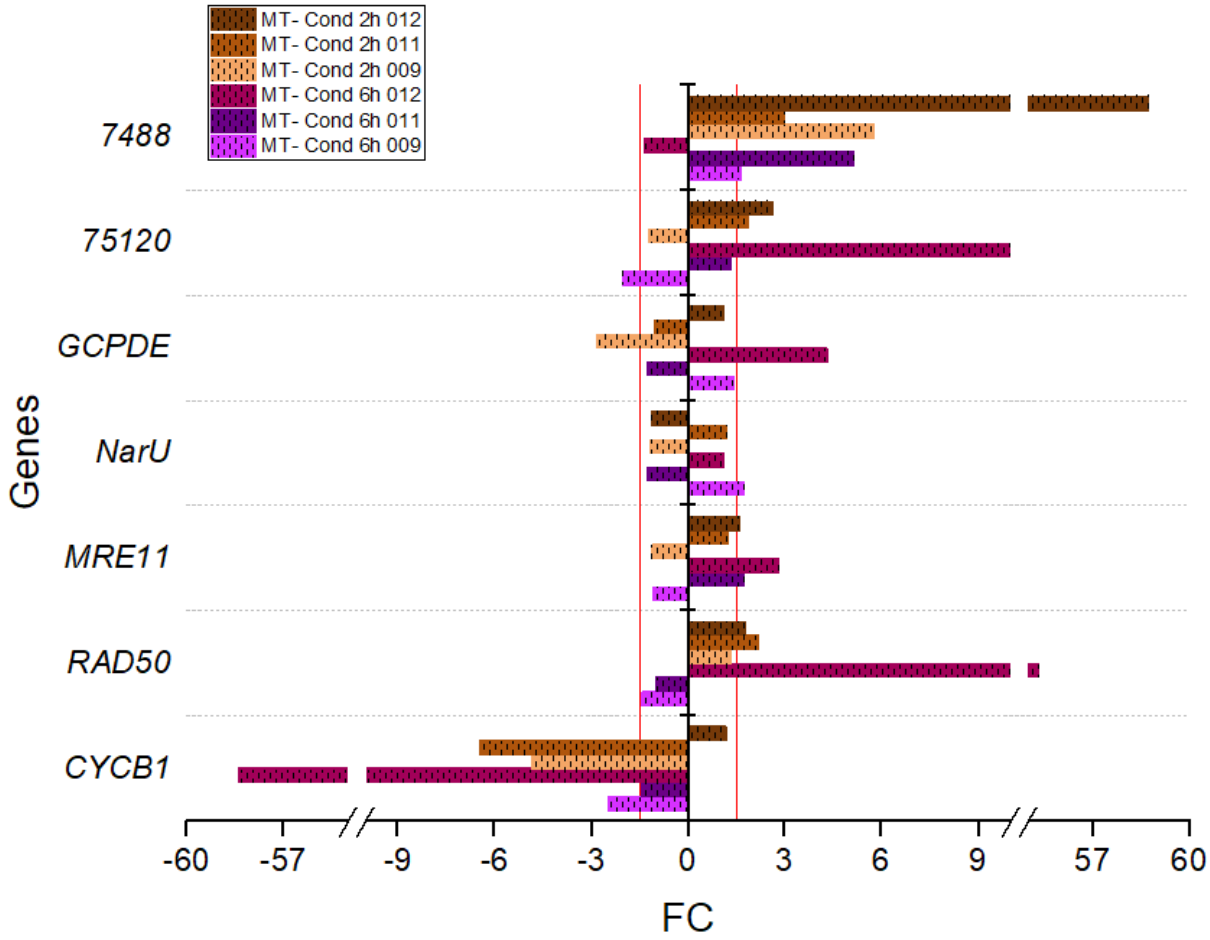


Fig.2.8 Fold change (FC) of the target genes in MT+ Conditioned samples with respect to the vegetative control. FCs are shown for the MT+ Cond at 2h and 6h, for all three experiments. The red line represents the threshold of FC equal to -1,5 and + 1.5.

Generally, I observed variability among replicates, time points and experimental condition. Table 2.4 allows to compare my results with those obtained by Basu et al. (2017): at 2 hours the general response in MT-S was more coherent than in MT+S, especially for *MRE11*, *NarU* and 75120; at 6 hours there was a higher concordance for all genes among experiments in MT+S, except for *NarU* and 7488.

Tab.2.4 Summary of the qPCR results for MT+ and MT- in the 'Sexualized' set-up and the RNAseq logFC from Basu et al. (2017). The direction of change for each gene is shown for all three experiments and for both time points. Orange and green indicate FC values that are below and above the significance threshold of -1.5 /+ 1.5, respectively. In white non-significant FCs.

		CYCB1	RAD50	MRE11	NarU	GCPDE	75120	7488	MRP1	CYCB1	RAD50	MRE11	NarU	GCPDE	75120	7488
		MT+ S								MT- S						
2h	9	∧	∨	∨	∧	∨	∨	∨	∧	∨	∨	∧	∧	∨	∧	∨
	11	∧	∨	∧	∧	∨	∧	∨	∨	∨	∧	∧	∧	∧	∧	∧
	12	∧	∨	∨	∧	∨	∨	∧	∧	∧	∧	∧	∧	∧	∧	∧
	Log FC Basu et al, 2017	**0,3**	**0,8**	**0,3**	**1,9**	**-0,6**	**0,07**	**0,16**	**-0,36**	**-0,5**	**-0,4**	**0,05**	**0,98**	**-0,16**	**0,04**	**7,8**
6h	9	∨	∧	∧	∧	∧	∧	∧	∧	∧	∨	∨	∨	∨	∨	∨
	11	∨	∧	∧	∧	∧	∧	∧	∧	∨	∧	∧	∧	∧	∧	∨
	12	∧	∧	∧	∧	∧	∧	∧	∧	∨	∧	∨	∨	∧	∨	∧
	Log FC Basu et al, 2017	**-1,9**	**0,8**	**1,3**	**-2,4**	**1,2**	**1**	**-0,25**	**3,6**	**-1,5**	**-0,4**	**0,8**	**-1,9**	**2,6**	**1,6**	**6,8**

In MT+ Cond the response of the investigated genes was very variable and only two genes showed a clear regulation: *GCPDE* was downregulated at both time points in two of three experiments, and 7488 was upregulated at 6 hours (Table 1.5). In the MT- Cond the general response was less variable: *CYCB1* was downregulated at both time points, *RAD 50* and *MRE11* were mainly upregulated, *75120* was upregulated in two experiments at both time points and 7488 was upregulated in all experiments except in one (Table 1.4). *NarU* and *GCPDE* did not display a clear regulation.

Tab.2.5 Summary of the qPCR results for MT+ and MT- in the conditioned medium experiments. The direction of change for each gene is shown for all three experiments and for both time points. Orange and green indicate FC values that are below and above the significance threshold of -1,5 /+ 1,5, respectively.

		CYCB1	RAD50	MRE11	NarU	GCPDE	75120	7488	MRP1	CYCB1	RAD50	MRE11	NarU	GCPDE	75120	7488
		MT+ Cond								MT - Cond						
2h	9	∧	∧	∨	∨	∧	∨	∨	∧	∨	∧	∨	∨	∨	∨	∧
	11	∨	∨	∨	∧	∨	∧	∨	∨	∨	∧	∧	∧	∨	∧	∧
	12	∧	∧	∨	∧	∨	∧	∧	∨	∧	∧	∧	∨	∧	∧	∧
6h	9	∨	∧	∧	∨	∧	∧	∧	∧	∨	∨	∨	∧	∧	∨	∧
	11	∧	∨	∨	∨	∨	∨	∧	∨	∨	∨	∧	∨	∨	∧	∧
	12	∨	∧	∨	∧	∨	∨	∧	∧	∨	∧	∧	∧	∧	∧	∨

Comparing the results of MT Cond with MT S for MT+, there were matches for regulation of *NarU* at both time points that displayed a different regulation respect to reference data, for 7488 at 6 hours that was upregulated; for *GCPDE* that was downregulated at 2 hours but then upregulated at 6 hours in MT+S while in MT+ Cond it was mainly downregulated. *MRP1*, known to be strongly induced at 6 hours in sexual reproduction (Basu et al.,2017), had an atypical behaviour in MT+ Cond at 2 and 6 hours showing an opposite trend respect to the MT+S.

In the results of MT Cond and of MT S in MT-, there was concordance in the regulation of *CYC1B* that was significantly downregulated in two out of three experiments; 75120 was upregulated at 2 hours (experiments 012 and 011); and, finally, 7488 was strongly affected by the filtered medium of MT+ showing upregulation with the exception of experiment 012 for the second time point. *RAD50*, *MRE11*, *NarU*, *GCPDE* trends were not consistent in MT- Cond.

2.3.4 Time course of gene expression in sexualized MTs

The time course experiments were carried out in the bipartite glass apparatus, with the two MTs growing in mono-culture as controls. This experiment aimed to document the expression dynamics of genes considered important during the early phase of sexual reproduction, specifically *MRM1*, *MRM2*, *MRP1*, *MRP2*, 7488, *NarU* and *CYCB1*. In Figure 2.9 the relative expression for *MRP1*, *MRP2*, *MRM1* and *MRM2* is shown; for *MRP1* the relative expression scale is different for the exp013 indeed the relative expression of this gene was already different in the controls (Figure 2.10) being highest in the exp013. In all three experiments, *MRP1* showed higher levels of transcripts in MT+ S respect to the control (MT+C) between 4 and 6 hours, and these levels peaked at 8 hours. Expression levels of *MRP2* and *MRM1* were very low, in MT+ and MT- S these two genes

were more expressed than in MT+ and - C in two out of three experiments at 8 hours. *MRM2* was induced starting from 4 h except in exp013 in which induction occurred later.

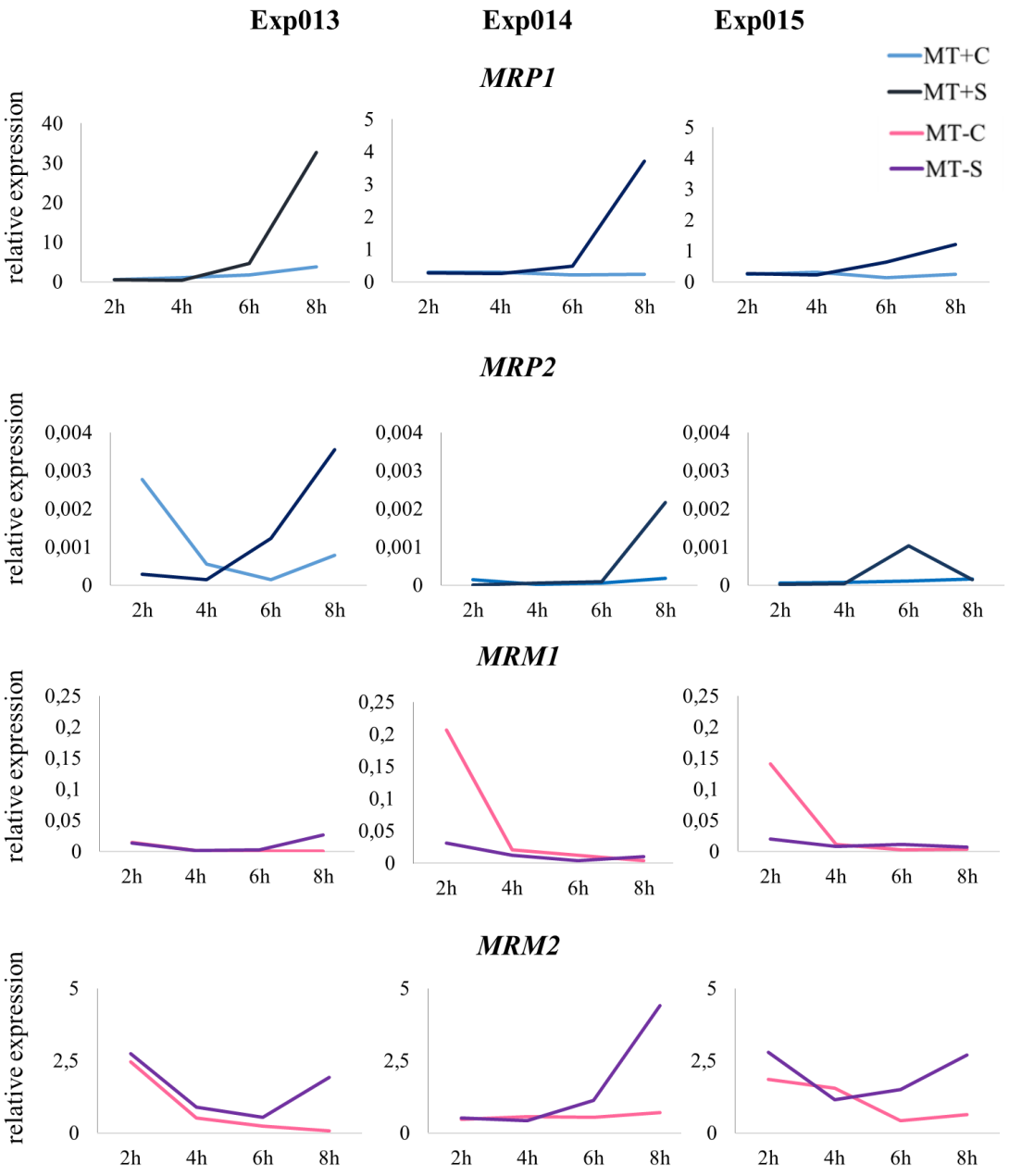


Fig.2.9 Expression profile of MR genes for three time course experiments (exp 013, 014 and 015). In dark blue and purple are represented the sexualized MT+ and MT– strains, respectively; in light blue and light pink are represented the MT+ and MT– controls cultures, respectively. On the y axis there are the relative expression values, on the x axis the time points.

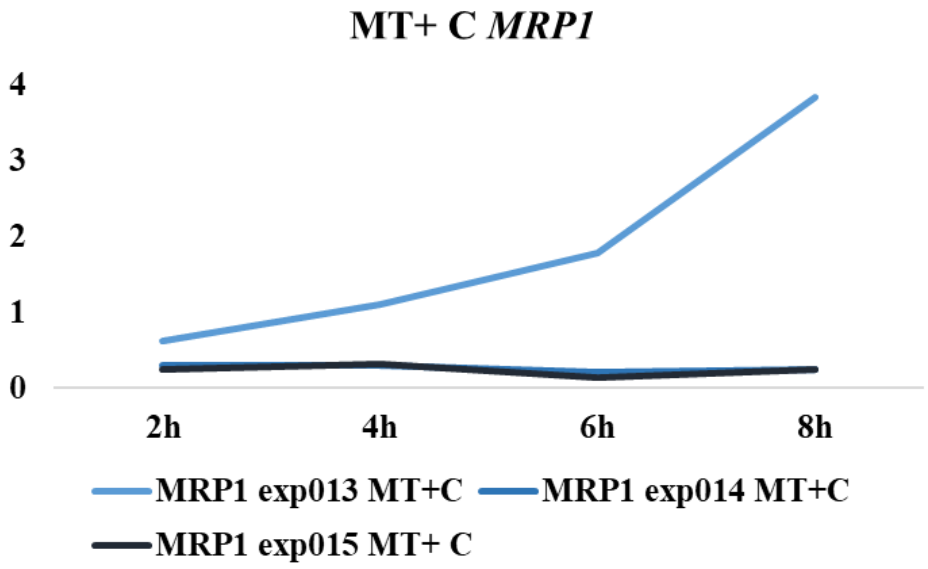


Fig.2.10 The relative expression profile of *MRP1* for all three experiments (exp 013, 014 and 015) during the time course in the control cells of MT+. On the y axis there are the relative expression values, on the x axis the time points.

Other investigated genes in the time course were *CYB1*, *NarU* and 7488 (Figure 2.11). Expression profiles of these genes were different in the three experiments, most likely due to the intrinsic variability of the sexual reproduction process. The only strong and consistent signal was the induction of 7488 in MT-S starting between 4 and 6 hours and becoming even more evident at 8 hours.

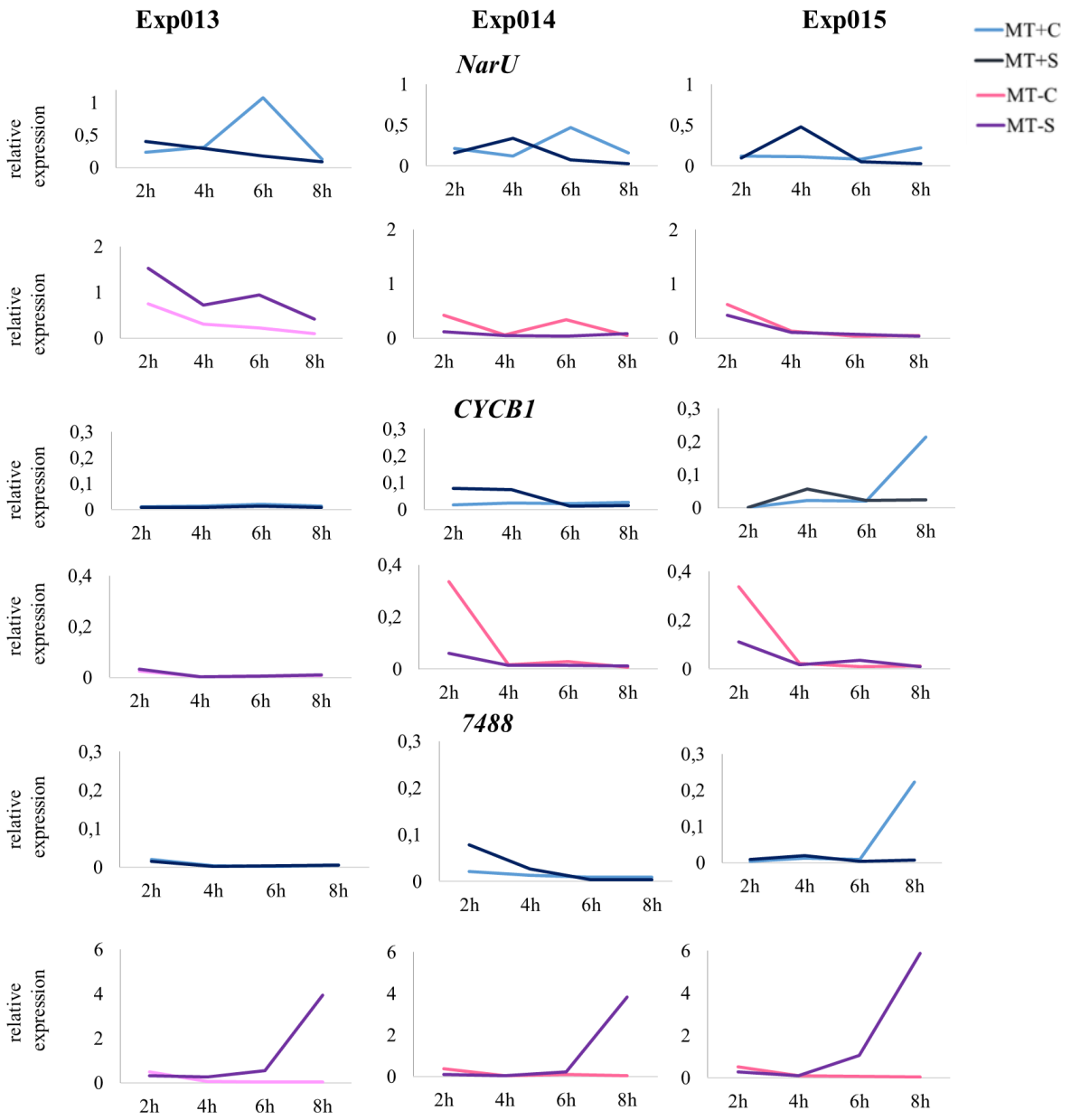


Fig.2.11 The relative expression profile of *NarU*, *CYCB1* and *7488* during the three time course experiments (exp 013, 014 and 015). In dark blue and purple are represented the sexualized MT+ and MT– strains, respectively. In light blue and light pink are represented the vegetative MT+ and MT– strains, respectively. On the y axis there are the relative expression values, on the x axis the time points.

2.3.5 Perturbation experiments with inhibitors

Perturbation experiments were performed to interfere with genes involved in putative signal transduction pathways. Chosen chemical compounds were: CPTIO, LY-83583 and TRIM. CPTIO is a scavenger of nitric oxide (NO), LY-83583 is an inhibitor of the soluble guanylate cyclase (sGC) and TRIM is an inhibitor of nitric oxide synthase (NOS). The genes followed in these experiments were *MRP1* and *sGC*. Both genes were upregulated in Basu et al. (2016), *MRP1* was hypothesized to be at downstream of the sGC signalling. The first step was to evaluate the effect of inhibitors on cells survival (toxicity tests) by microscope examination at 6 and 48 hours after addition of the compound. After the concentration of inhibitors to use had been determined, experiments to study gene expression changes in the presence of inhibitors were performed. Toxicity tests were also performed on co-cultured MT+ and MT- to check the non-lethal maximum dose of the inhibitor at which sexual reproduction was affected. The concentrations for CPTIO tests were 10, 5 and 1 µM. CPTIO did not affect cell survival and the efficiency of sexual reproduction at any of the concentration tested (Table 2.6).

Tab.2.6. Treatment with CPTIO on MT -, MT+ and cross. Effects evaluated at 6 and 48 hours. In the cross the effect of CPTIO on sexual reproduction was evaluated.

	Effect after 6 hours			
	10 µM	**5 µM**	**1 µM**	**Control**
MT-	alive	alive	alive	alive
MT+	alive	alive	alive	alive
Cross	alive	alive	alive	alive
	Effect after 48 hours			
	10 µM	**5 µM**	**1 µM**	**Control**
MT-	alive	alive	alive	alive
MT+	alive	alive	alive	alive
Cross	Initial cells	Initial cells	Initial cells	Initial cells

The LY-83583 concentrations 2, 1, 0.2, 0.1, 0.01 µM were lethal after 6 hours of treatments, while 0.001 and 0.0005 µM were not lethal; these concentrations delayed gamete production and auxospore formation with respect to control (Table 2.7).

Tab.2.7. Treatment with LY-83583 on MT -, MT+ and cross. Effects were evaluated at 6, 48 and 96 hours. In the cross the effect of LY-83583 on sexual reproduction was evaluated. "Control" refers to the samples treated with the solvent only.

	Effect after 6 hours				
	2 µM	**1 µM**	**0.2 µM**	**0.1 µM**	**Control**
MT-	dead	dead	dead	dead	alive
MT+	dead	dead	dead	dead	alive
Cross	dead	dead	dead	dead	alive
	Effect after 6 hours				
	0.01 µM	**0.001 µM**	**0.0005 µM**	**Control**	
MT-	dead	alive	alive	alive	
MT+	dead	alive	alive	alive	
Cross	dead	alive	alive	alive	
	Effect after 48 hours				
MT-		alive	alive	alive	
MT+		alive	alive	alive	
Cross		gametes	gametes	initial cells	
	Effect after 96 hours				
MT-		alive	alive	alive	
MT+		alive	alive	alive	
Cross		Auxospores	Auxospores	Initial cells	

For TRIM the tested concentrations were 1, 0.1 and 0.01 mM. While the 1 mM TRIM was lethal, 0.1 and 0.01 mM did not affect cell survival (Table 2.8). The cross treated with 0.1 mM did not show sexual stages.

Table 2.8. Treatment with TRIM on MT -, MT+ and cross. The effects were evaluated at 6 and 48 hours. "Control" refers to the samples treated with the solvent only.

	Effect after 6 hours			
	1 mM	**0.1 mM**	**0.01 mM**	**Control**
MT-	dead	alive	alive	alive
MT+	dead	alive	alive	alive
Cross	dead	alive	alive	alive
	Effect after 48 hours			
		0.1 mM	**0.01 mM**	**Control**
MT-		alive	alive	alive
MT+		alive	alive	alive
Cross		No Initial cells	Initial cells	Initial cells

Since CPTIO did not affect sexual reproduction it was not used in perturbation experiments, for LY-8358 the results were not replicable (delay in sexual stages formation was not consistently observed), so gene expression changes were evaluated only in the treatment with TRIM 0.1 mM. The experiments were setup in multi-wells with insert membranes to keep MT + and MT- separated while allowing chemical communication. Samples were collected after 6 hours of treatment. RNA yields for MT- samples were very low and did not allow to perform analyses, only MT+ treated and non-treated samples were processed for qPCR analysis for the two target genes *MRP1* and *sGC*. The fold changes (FC) are shown with respect to control cells, i.e., the individual MT in monoculture (Figure 2.12). In MT+S treated with solvent (positive control) *sGC* was not regulated, while *MRP1* was upregulated. In MT+S treated with TRIM, both genes were upregulated. The perturbation experiment was replicated 4 times but sexual reproduction was not affected, there was no arrest of gametes production.

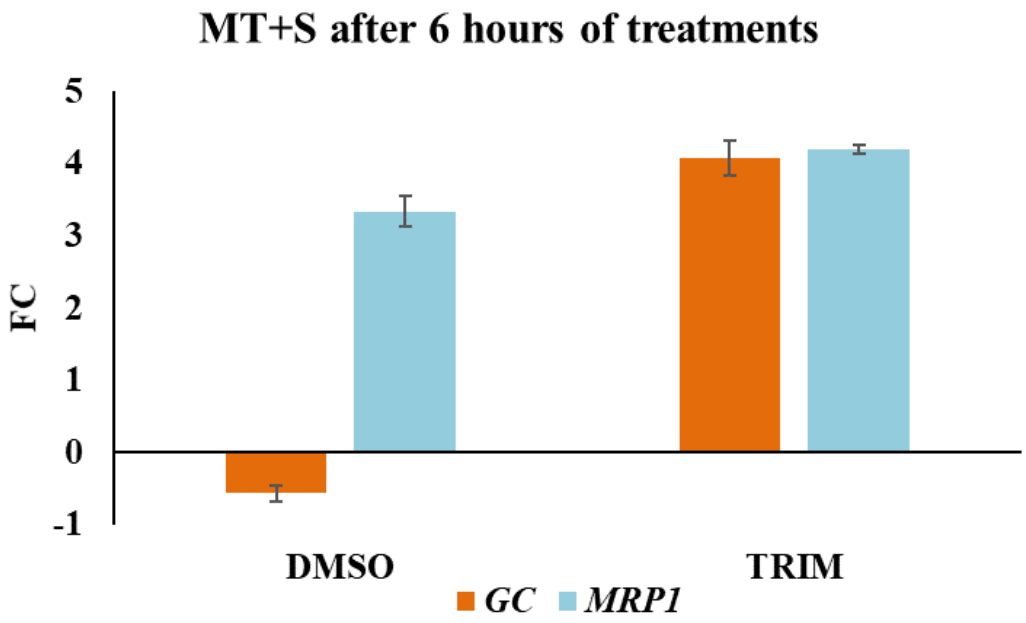


Fig.2.12. Fold changes of expression of *sGC* and *MRP1* in MT+ sexualized with respect to MT+ non-sexualized in DMSO and in TRIM after 6 hours. Error bars represent the standard deviation of technical replicates.

2.3.6 Sex ratio experiments

The aim of this experiment was to understand whether unbalanced sex ratios can affect mating efficiency. The experiments were carried out with three different couples (B, C and D) with different size for a total of six experiments (Table 2.1). After 24 hours from the set-up of co-cultures the mating events (gametangial pairs and cell clusters) were counted (Fig. 2.13-2.14). In all experiments sexual reproduction, in terms of number of mating events, was most favoured in the MT-:MT+ sex ratio equal to 50:50, followed by MT-:MT+ sex ratio equal to 10:90. Ratio MT-:MT+ equal to 90:10 showed the lowest number of mating events.

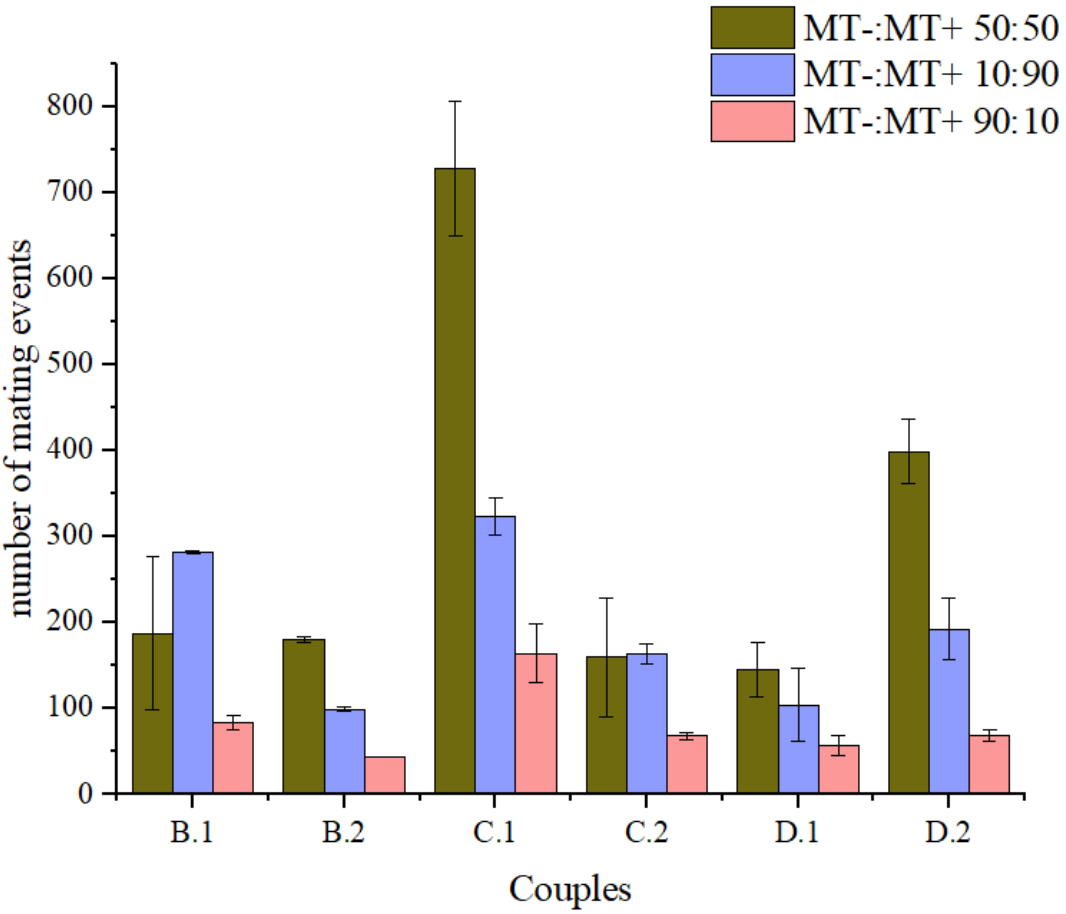


Fig.2.13. For each couple, the number of mating events for each sex ratio is shown (MT-:MT+ 50:50 in green, MT-:MT+ 10:90 in purple and MT-:MT+ 90:10 in pink). Average values ± st. dev. (n=3). The experiment was performed twice for each couples.

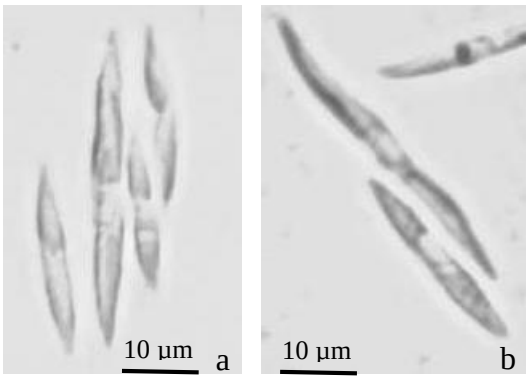


Fig.2.14. Exemple of mating events in the co-culture after 24 hours, a : cluster with one MT- and three MT+ cells b: pair of two opposite MTs.

2.3.7 Target genes expression profiles during later stages of the sexual reproduction process

To have a more comprehensive view of the expression profile of target genes considered in the studies illustrated above, I explored RNAseq data generated from an experiment aimed to elucidate the transcriptional changes occurring in the cross (MT+ and MT- cells in co-culture) at three time points (1 hour, 24 hours and 5 days).

In Table 2.9 the FC of genes in the following comparisons are shown:

- Cross 24 hours versus Cross 1 hour
- Cross 5 days versus Cross 1 hour
- Cross 5 days versus Cross 24 hours
- Cross 24 hours versus MT+ parental strain
- Cross 24 hours versus MT- parental strain
- Cross 5 days versus MT+ parental strain
- Cross 5 days versus MT- parental strain

Tab.2.9. FCs of target genes in several comparisons between controls and crosses. The expression levels obtained in the cross after 24 hours were compared with those of the controls (MT+ and MT-) and of the cross after 1 hour. The expression levels obtained in the cross after 5 days were compared with those of the controls (MT+ and MT-) and of the cross after 1 hour and after 24 hours. NC= no change. Genes are grouped based on their putative involvement in specific pathways or mechanisms (like in tab 3): in peach MR genes, in grey nutrient uptake, in pink genes involved in signalling, in orange unknown genes, in light blue genes related cell cycle.

	Cross 24h vs			**Cross 5 days vs**			
Gene name	**MT+**	**MT-**	**Cross 1h**	**MT+**	**MT-**	**Cross 1h**	**Cross 24h**
Nar U	-2,89	-3,57	-3,08	-2,69	-3,36	-2,87	NC
75120	1,87	1,73	NC	1,37	1,23	NC	-0,5
GCPDE	2,02	NC	0,99	3,18	0,95	2,14	1,15
sGC	6,53	7,95	NC	6,62	8,04	NC	NC
Adwaark	5,28	5,26	-0,62	5,53	5,5	NC	NC
Cathepsin	8,52	4,46	-0,69	8,22	4,16	-0,99	NC
CYCB2/2	1,38	1,7	1,28	NC	0,74	NC	-0,96
CYCB1	NC	NC	2,38	-1,29	-1,23	NC	-1,86
MRE11	0,99	0,86	NC	0,6	0,48	NC	-0,38
RAD50	1,39	1,38	1,48	0,44	0,44	0,54	-0,94
RAD51	7,84	7,92	1,73	5,25	5,33	-0,86	-2,59
7488	7,03	3,86	-0,97	6,64	3,47	-1,36	NC
31872	7,95	7,88	1,69	6,27	6,2	NC	-1,68
MRP1	5,87	8,56	3,63	4,3	6,98	2,05	-1,58
MRP2	1,45	5,64	-0,94	2,52	6,7	NC	1,07
MRM1	8,04	0,72	-1,32	8,06	0,75	-1,29	NC
MRM2	8,76	0,91	NC	8,04	NC	-1,06	-0,72

With a different visualization of the data (Figures 2.15 and 2.16) showing counts per milions (CPM) for vegetative cells and crosses (1 hour, 24h and 5 days) it was possible to appreciate the difference in basal gene expression levels among samples and to undestand more easily the trend of genes expression over time.

NarU decresed its expression in crosses at 24 hours and 5 days, while *75120, GCPDE, sGC, aardvark* and *cathepsin* incresed their expression in crosses at 1 hour, 24 hours and 5 days and among these genes that showed different trends, *sGC* mantained its expression levels at all time points (Figure 2.15.a). Expression levels of *CYCB2/2*, *CYCB1*, *MRE11*, *RAD50*, *RAD51-A1* peaked at 24 hours (Figure 2.15.b). 7488 peaked at 1 hour and its

expression decresed at 24 hours and 5 days, while 31872 peaked at 24 hours (Figure 2.16.a). The four MR genes considered are all induced in crosses, *MRP1* being again the one displaying the highest levels of expression. *MRP2* had higher expression levels at 1 hour and 5 days. *MRM1* increased only moderately at 1 hour, decreased its expression at 24 hours remainig constant at 5 days. *MRM2* was strongly upregulated in crosses at 1 hour and 24 hours, decreasing slightly at 5 days. Note that the MR genes and *7488* had different levels of expression in MT+ compared to MT-, this was more evident when looking at expression levels (Figures 2.15 and 2.16) rather than at the FCs. For instance, *MRM1* and *MRM2* were strongly upregulated in crosses versus MT+, this happens because they both have lower basal expression levels in vegetative MT+ strains (Vitale's thesis, Basu et al. 2017, Russo et al. 2018). *MRP1, MRP2, MRM1* and *MRM2* showed different expression trends over time: *MRP1* peaked at 24 hours, *MRP2* at 5 days while *MRM1* and *MRM2* peaked at 1 hour; among these *MRP1* was the gene with highest expression levels in the crosses (Figure 2.16.b).

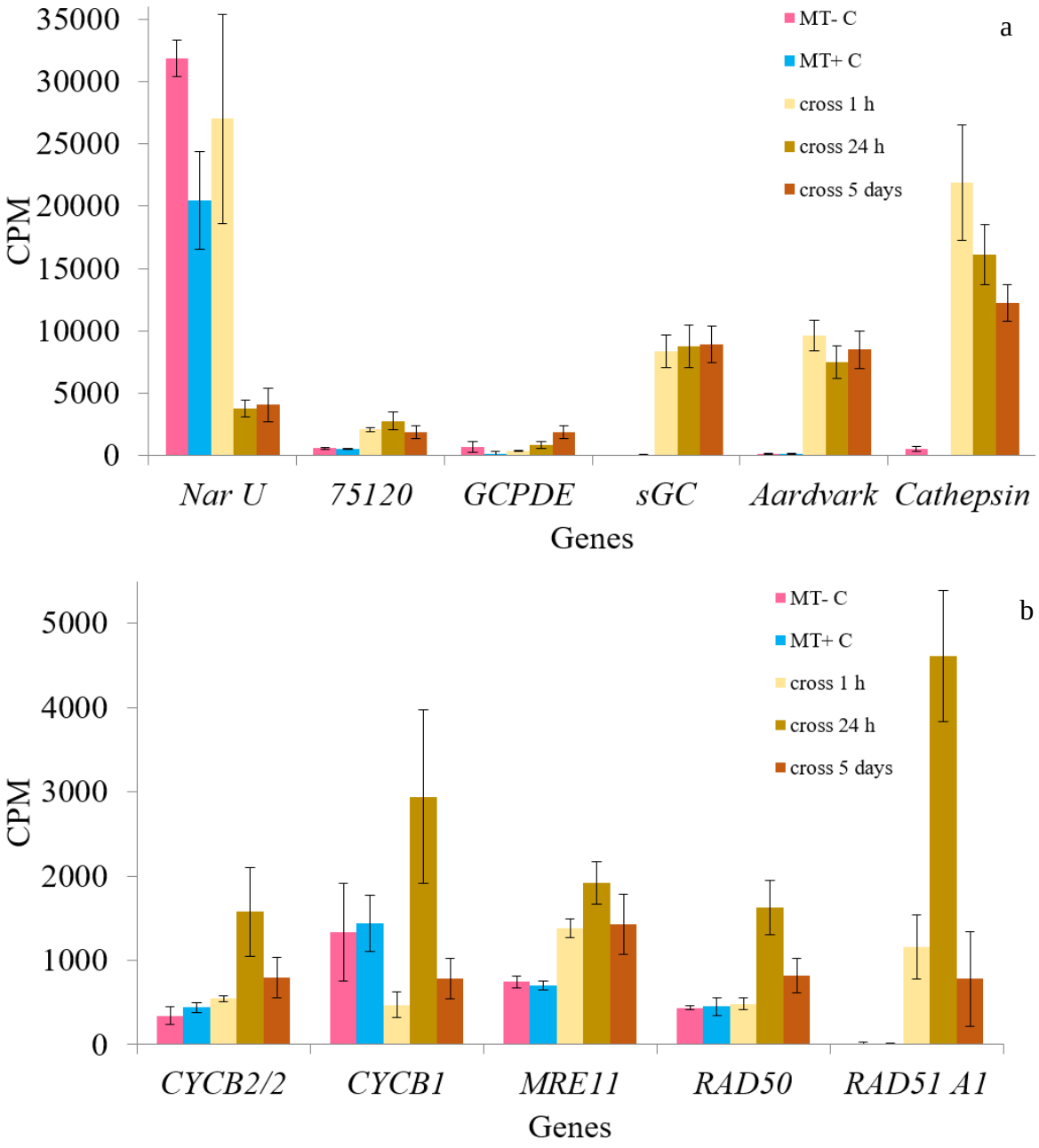


Fig.2.15. The read counts in counts per million (CPM) of vegetative cells and crosses; a: *NarU, 75120, GCPDE, sGC, aadvark and cathepsin*; b: *CYCB2/2, CYCB1, MRE11, RAD50* and *RAD51-A1*. Error bars represent standard deviations of three biological replicates.

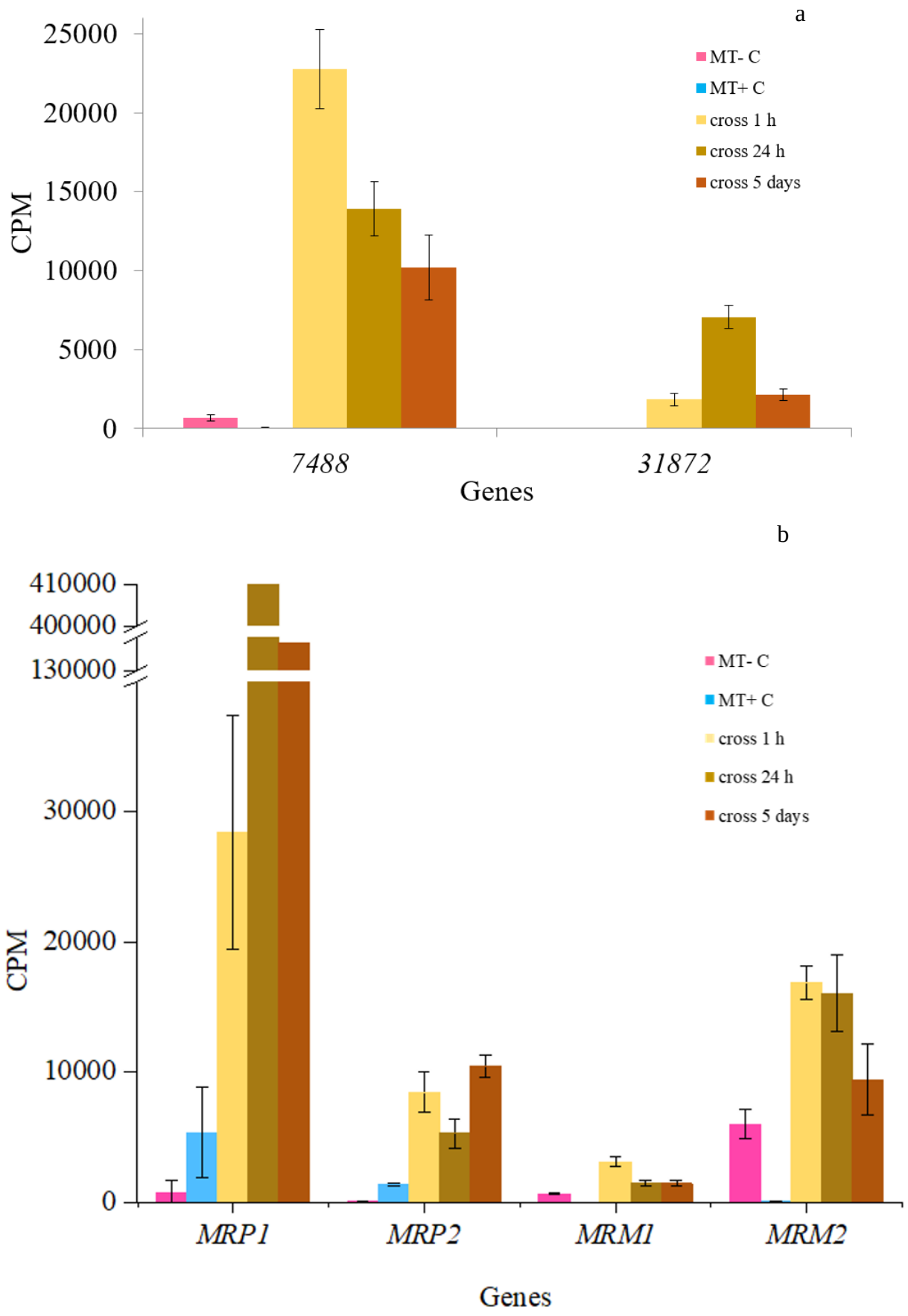


Fig.2.16. The read counts in counts per million (CPM) of vegetative cells and crosses; a: *7488* and *31872*; b: *MRP1*, *MRP2*, *MRM1* and *MRM2*. Error bars represent standard deviations of three biological replicates.

2.4 Discussion

The experiments presented in this chapter throw light on a part of the gene machinery of sexual reproduction, highlighting important aspects in the dynamics of action of the genes investigated. Furthermore, comparing these results with the known reproduction system of *S. robusta* will help to understand the differences between planktonic and benthic species with respect to mate perception. The effort to compare these data presents a limit linked to the different experimental set-ups used: opposite MTs in chemical contact; MTs treated with filtered medium and opposite MTs in co-culture. These three different set-ups affected timing of the process that it is already variable when different strains are used. Nevertheless, merging of all these results provided novel information about the investigated genes, that are discussed as follows.

2.4.1 Identification of a molecular marker gene to isolate putative pheromones in the MT+ conditioned medium

Using the medium conditioned by the opposite MT to induce sexualisation in MT+ and MT- was a strategy chosen to identify a marker gene for isolating the *P. multistriata* pheromones. In *S. robusta* the first identified pheromone was diproline (Gillard et al. 2013). Starting from the observation that the MT+ increased motility in the presence of the medium of MT-, a study for the identification of the chemical cue inducing motility was carried out using cartriges loaded with different fractions of the medium. Diproline was characterized using a metabolomics approach applied to the fraction that elicited a positive response (Gillard et al. 2013). Unfortunately, in *P. multistriata* a unidirectional attracting behaviour (one MT being strongly attracted by the opposite MT) to be used for the identification of putative pheromones was never observed. Investigating gene expression changes in cells treated with conditioned medium was thus a possible way to obtain a potential bioassay.

The sexualized MT+ cells (MT+ S) of all three independent experiments at 6 hours showed for all genes, except for *CYCB1*, concordant behaviour in expression changes. *MRP1*, a gene with expression levels normally already higher in vegetative MT+ than in vegetative MT-, was always upregulated in MT+S in all experiments at 6 hours, and this gene would be a good candidate to be used as a proxy for the onset of sexualisation when MT+ cells perceive the presence of MT-. However, *MRP1* in MT+ Cond was not always induced, perhaps due to a high variability in the strength of signals emitted and in the timing of the process. Alternatively, the MT- pheromone stimulating MT+ could be unstable in the medium, so in some cases the filtered medium of MT- would be unable to induce *MRP1* and the other genes involved. Adding new filtered medium of MT- vegetative cells could be tried to ensure MT- pheromone supply. Since the activation observed in MT Cond 6 hours in exp 009 is not very strong (Fig.2.12), it is also possible that for a strong and sustained *MRP1* induction the presence of a second MT- pheromone, secreted only after the MT- has been sexualized, is necessary.

Considering all the data together, it is possible to imagine a scenario in which *MPR1* is an MT+ primary signal constitutively expressed at low levels. When the MT+ perceives MT- cues it upregulates *MRP1* expression to strengthen the message, signalling its presence, to further stimulate the opposite MT with higher level of MRP1. *MRP1* contains in its protein sequence a putative signal peptide that targets proteins for translocation across the endoplasmic reticulum (ER) membrane and possibly to the extracellular space (Russo et al., 2018), it could therefore be an MT+ pheromone itself. Further experiments will be needed to clarify if *MRP1* induction is due to a primary (from the MT- vegetative) or to a secondary (from the MT- sexualized) pheromone.

In any case, from what I observed, *MRP1*, despite being one of the genes displaying the strongest response to sexualisation (see also Fig.2.22), does not appear to be the best gene to follow to isolate the MT- pheromone.

In the sexualized MT- cells (MT- S) the expression of target genes is more fluctuating (Table 2.4); these different general responses could reflect the fact that the time windows of mate perception and then also of sexual reproduction are depending on the physiological state of the cells. Moreover, in addition to expressing a set of MT-linked genes characterising uniquely the response of either one or the other MT, the two MTs also upregulate common genes but can, at times, do this with a different timing. In *S. robusta* MT- produces diproline after perception of SIP+ and then diproline is perceived by MT+, before this step SIP- could induce the increment of SIP+.

Considering the expression changes in MTs treated with the filtered medium of opposite MTs (MTs Cond), it is important keep in mind the fact that the sexualized treatments and treatments with filtered medium are different in terms of cue composition: in the sexualized treatment there are all cues that opposite MTs produce and exchange between them, in other words there are all cues that are produced during the cross talk; in the treatment with filtered medium of opposite MTs the putative cues produced after the mate perception are absent. Furthermore, in the sexualisation treatments all cues are diluted two times, while in the filtered medium treatments they are not diluted. Nonetheless, many of the genes responded to the filtered medium and sexualisation in a concordant manner. Comparing the MT+ S and MT+ Cond at 6 hours, the gene *7488* was upregulated, even if for two different experiments the FC are slightly below the threshold (Tables 2.4 and 2.5). *7488* in the MT+ Cond 6 hours exp012 did not pass the threshold, but at 2 hours in the same experiment FC was equal to 2.8, this could be related to variability in the time windows of mate perception and to the fact that in this experiment the general response to sexualisation was very strong (see figures 2.11 and 2.12). The gene *7488* in the MT- Cond was upregulated in all experiments at 2 hours, while at 6h in two out of three experiments. This gene is definitely involved in both MTs in the response to sexual cues (see also Fig.2.22), responding first in MT- and then in MT+. 7488 is therefore a good candidate

marker to use for the determination of the medium fraction containing the pheromones. Since MT- responds more consistently (Table 2.5) and possibly earlier than the MT+ (see also Table 2.4 and Basu et al., 2017), Future experiments would be useful to try to isolate the MT+ cue that triggers response in the MT-.

It is also interesting to note that 7488 levels in vegetative MT- seem to be higher than in vegetative MT+, based on this difference 7488 would behave like MR genes and could indeed be considered a sixth MR gene, specific for MT- (*MRM3*). This result should be confirmed using a higher number of independent MT+ and MT- strains.

The gene *CYCB1* in the MT- S at 6h in two experiments was downregulated. In the MT- Cond at 2 hours in two experiments *CYCB1* was downregulated, at 6 hours in all three experiments this gene was downregulated. In *S. robusta*, the expression of *CYCB1* was repressed in MT- treated with SIP+ (Moeys et al. 2016). Furthermore, this gene was highly expressed during mitosis in the diatom *Phaeodactylum tricornutum* (Huysman et al. 2010). These results suggest that *P. multistriata* MT- treated with medium of MT+ cells could have perceived the presence of the MT+ responding also with a gene involved in cell cycle regulation. This result confirmed the effect of medium on MT-, similarly to *S. robusta* in which the treatment with SIP+ was validated also by downregulation of *CYCB1*. So in both species the perception of MT+ cues arrest the cell cycle.

At 2 hours in both MT- S and Cond, 75120 was upregulated in two experiments. This gene was upregulated in the RNAseq data during sexualisation (Basu et al., 2017, see Supplementary_table_09). The gene 75120 is annotated like Type II inositol 1 4 5-trisphosphate 5-phosphatase. In response to extracellular signals the inositol triphosphate (IP3) can be used as second messenger. IP3 is produced, by cleaving of phosphatidylinositide (4,5) biphosphate (PIP2), in the plasma membrane layer and then diffuses in the cytosol to bind to intracellular receptors that release Ca2+ (Alberts 2015).

In plants the abscisic acid pheromone is related to the IP3 pathway (Gilroy et al. 1990; Burnette 2003). In eukaryotic cells the phosphoinositides play a key role in membrane trafficking of the vacuole and this role was firstly demonstrated in yeast (Pfeffer 1996). The inositol ring of phosphatidylinositol can be reversibly phosphorylated at one or a combination of positions (3', 4' or 5') resulting in seven different second messengers involved in cellular processes like growth, differentiation, cytoskeletal rearrangements and membrane trafficking (Martin 1998). This signal of these second messengers is terminated by 5-phosphatases that remove the phosphate from the inositol rings. A role of this phosphatase was elucidated in vesicular transport by synaptojanin, a protein in mammalian nerve cells that is involved in synaptic-vesicle endocytosis (McPherson et al. 1996), and these synaptojanin proteins are also implicated in yeast endocytic membrane traffic (Singer-Krüger et al. 1998). In the MT- S and Cond at 2 hours the upregulation of the Type II inositol 1 4 5-trisphosphate 5-phosphatase (75129) could be related to one of the many steps of mate perception, such as a modulation of the initial phase of mate recognition by another second messengers (Martin 1998) or interruption of IP3 signal (Berdy et al. 2001). In Basu et al., the gene PSNMU-V1.4_AUG-EV-PASAV3_0072880.1 annotated like G protein-coupled receptor (GPCR) is upregulated. GPCRs trigger the downstream inositol phospholipid signalling pathway, these receptors are involved in mating of yeast (Alvaro &Thorner, 2016). 75120 in MT+S at 6 hours, like in the RNAseq data (Basu et al. 2017), was upregulated, however this signal was not present in the MT+ Cond. This data indicate that 75120 activation is secondary to another cue that derives from MT- after it has become sexualized and that therefore is not present in the medium of vegetative MT-.

These data allow me to put 75120 downstream in the cascade of genes required for sexualisation, indicating that GPCRs and the signalling pathway illustrated above could respond to secondary and not to primary cues.

CGPDE in MT+ S was downregulated at 2 hours and upregulated at 6 hours; in the MT+ Cond at both time points it was repressed except in one sample where it was upregulated. In *S. robusta* the expression of this gene peaks at 3 hours after adding SIP+, indicating that it could be an important gene for perception of SIP+ and for the intercellular signal transduction of it. In *P. multistriata* this gene was also implicated in perception but, similarly to 75120, it could not be among the primary players of the first perception step.

From these results it can be concluded that:

1. Even when the mate perception phase is proceeding successfully, as testified by the presence of sexual stages on the day following the experiment, the molecular response can be variable in terms of time and strength. This variability led to a noisy transcriptomic response in different experiments. Nevertheless, signals emerged and allowed a preliminary reconstruction of a hierarchy of gene action.

2. Lack of strong and consistent induction of *MRP1* in MT+ Cond suggests that a second cue from MT- might be needed after MT- is sexualized, and therefore that *MRP1* is not the best candidate in a bioassay to isolate the primary MT- pheromone.

3. The mate perception is articulated in multiple steps and it is an asymmetric process, the strongest response in treatments with filtered medium is in MT- Cond showing strong regulation of the 7488 and *CYCB1* genes at both time points.

4. The gene 7488 can be used as molecular marker to test different medium fractions of MT+ on MT- to isolate the first putative pheromone.

2.4.2 Timing gene activation in the early phases of sexual reproduction

The time course was carried out on a subset of genes investigated (*NarU*, *CYCB1* and 7488) in the experiments with filtered medium plus the following additional genes: *MRP1*, *MRP2*, *MRM1* and *MRM2*. The aim was to elucidate the sequential activation of these

genes during mate perception. 7488 is a candidate molecular marker for testing MT+ medium fractions on MT- since in the MT-Cond this gene was induced, and the aim was to investigate its trend during sexualisation in both MTs. *NarU* and *CYCB1* were chosen as representative genes to define the timing of the downregulation of nutrient uptake and of the cell cycle arrest, respectively. *MRP2* and *MRM2* are putative receptors and *MRM1* is possibly a transcription factor and information about their response during sexualisation might yield information about the relationship between these and other potential key genes.

The *MRP1* gene in MT+S started to increase its expression from 4-6 hours, and its expression was the highest at 8 hours. The expression trend of *MRP2* was not very consistent in all three experiments, although there was a tendency to have higher expression at 6 (exp 013 and 015) and 8 hours (exp 013 and 014). Moreover, its basal expression levels were always very low. As for *MRP2*, the expression levels of *MRM1* were very low, and also the trends were not very clear. The expression levels of *MRM2* were higher with respect to *MRM1*. For *MRM2* the trend was more clear, in exp013 its expression increased from 6-8 hours, while in exp014 and015 the increase in expression was anticipated at 4-6 hours.

The levels of *CYCB1* and *NarU* did not change dramatically in any of the two MTs, even if there was a declining trend in both sexualized MTs. It is possible that given the high time variability of the process and of the biological system some genes are less stable in their expression profile. 7488 showed a gradual increase only in MT-, confirming a clear signal in MT-. This gene started to rise from 4 hours and its expression increased at each time point. Regarding the MR genes, an interesting observation is that the basal expression of *MRP1* in the vegetative cells was different in the three experiments, and that the response of *MRPs* and *MRM1*, in sexualized MTs, was weak or absent when the basal expression of *MRP1* in the vegetative MT+ was low. Despite all three experiments were

performed under the same experimental condition, *MRP1* in control cells has the highest expression level in exp 013, while in exp 014 and 015 the expression levels are comparable (Figure 2.16). The circadian expression trend of this gene in MT+ was examined in a PhD thesis (Vitale PhD thesis, 2015), in this experiment it was highlighted that in cells synchronized with 36 hours of dark the expression level of *MRP1*, after 2 hours of re-illumination, was very low, increasing for the next two hours and then remaining stable over the next 24 hours. It is known that prolonged darkness can affect molecular pathways involved in several processes that in this condition have a free-running rhythm becoming out of normal phase (Foster 2017). Darkness leads to a downregulation of *MRP1* expression, the time required to restore stable levels of this gene after exit from the dark might be variable, and might account for part of the variability that I saw in the other genes.

In the green alga *Chlamydomonas reinhardtii* three genes linked to specific MT loci were identified: *FUS1* on the MT+ locus, *MTD1* and *MID* on the MT- locus (Sekimoto 2017). *MID* is expressed at basal level in MT-, and during sexual reproduction (triggered by nitrogen depletion) it is induced after 30 minutes and its expression level returns to basal expression at 1 hour and finally at 4-6 hours its expression is strongly induced (Lin and Goodenough 2006). The first induction of *MID* induces the expression of *MTD1*. The second induction of *MID* is triggered both by nitrogen starvation and *MTD1* but only the level reached after the induction of *MTD1* is enough to activate minus gamete-specific genes required for differentiation of MT- vegetative cells in gametes (Lin and Goodenough 2006). The induction levels in MT- S of *MRM2* are different among experiments. When the basal expression of *MRP1* in MT+ C cells is lower (exp 014 and 015), the increase in expression of *MRM2* in sexualized cells occurs between 4 and 6 hours, while in exp 013, when the *MRP1* basal levels in controls are higher, it occurs between 6 and 8 hours. This correlation could be only dependent on the variability of our

experimental system, or it could be due to the way the system works. In other words, it seems that the induction of *MRM2* is anticipated when the expression level of *MRP1* is weaker. To verify this correlation between the levels of *MRP1* and the MT- response, time course experiments could be performed investigating the response of MR genes with the different cell densities of the MT+ culture. The sexual reproduction success in *P. multistriata* seems indeed to be correlated with cell density (Scalco et al. 2014).

2.4.3 The putative involvement of NO as second messenger in mate perception

Although the perturbation experiments with inhibitors were not replicated, they allowed to infer useful indications about putative pathways involved. The NO radical scavenger CPTIO did not affect sexual reproduction despite the tested concentrations were high. This could mean that NO is not implicated in mate perception or that the available NO is not limiting for the biological process under investigation. LY-83583 is an inhibitor of the guanylate cyclases. The sGC gene in Basu et al. was upregulated in both sexualized MTs. The tested concentration of LY-83583 did not block sexual reproduction, but anyway a delay in gamete production was observed. The gametes normally formed zygotes, auxospores and initial cells indicating that the compound affected parental cells, perhaps causing a delay in the mate perception phase, rather than interfering with the production of viable and functional gametes. The NOS inhibitor 0.1 mM TRIM in some experiments blocked sexual reproduction. This effect was investigated performing qPCR in MT+ S treated and control on two target genes, *MRP1* and sGC. The hypothesis was that if NOS or NO are involved in mate perception, the sGC, that is activated by NO, would not produce cGMP. Moreover, if the *MRP1* induction is a response downstream of sGC activity its expression would be repressed. In MT+S treated with the solvent (control) *sGC* was, unexpectedly, not regulated and *MRP1* was upregulated; in MT+S treated with TRIM both genes were upregulated. This could mean that there is a negative feedback on the *sGC* transcription: in cells treated with TRIM, NO limitation resulting in impaired sGC

activation would be sensed by the cell that tries to compensate producing more sGC transcript. Unfortunately, it was not possible to replicate the results for technical problems related to quantity and quality of RNA and to phenotype variability. The *MRP1* induction seemed however to be independent from the sGC. Data therefore support the hypothesis of a very early role for *MRP1*, most likely upstream of the other genes investigated in this study. Although the data produced so far are not conclusive, they provide an indication that NO could actually be involved in the process but additional experiments will be needed to test this hypothesis. In literature there are many evidences of NO involvement in cell fertilization, in sea urchin NOS is fundamental for egg activation and necessary and for successful fertilization (Kuo et al. 2000). In plants NO is involved in many important processes like development and signalling in biotic and abiotic responses (Domingos et al. 2015). Furthermore, NO is important also for cell–cell communication during sexual reproduction, namely in pollen tube guidance (Prado et al. 2008), and in the self-incompatible system that is important for plant to avoid inbreeding and promote outcrossing (Wilkins et al. 2011). In plants, even if NO represents an important messenger, the NOS gene is not present; the production of NO occurs via a NAD(P)H-dependent nitrate reductase, a cytosolic enzyme associated with nitrogen assimilation, or by a mitochondrial electron transport-dependent reductase, using arginine as a substrate, like animal NOS (Domingos et al. 2015). There is no evidence of NO involvement in diatom reproduction and in mate perception, but as evident in this thesis the studies in this group of organisms are very limited. NO passes across membranes and diffuses very rapidly; these features make NO an ideal molecular signal for cell-cell communication in a process that has to be synchronized and that requires rapid responses.

2.4.4 Asymmetry in the formation of mating events in unbalanced sex ratios

Investigating the dynamic of pair and cluster formation helped to understand the attractive phenotypes in *S. robusta*. In this species a migrating mating type (MT+) moves towards an

attracting cell (MT-) forming pairs or clusters around the MT-. In *S. robusta*, setting different sex ratios it was observed that in 50:50 and in 10:90 MT+:MT- ratios there was mainly formation of pairs, while clusters could be observed with a 10:90 MT-:MT+ ratio, because a few MT- cells were able to attract many MT+ cells (Gillard et al. 2013). As mentioned above, a clear attraction phenotype cannot be observed in *P. multistriata,* however similar experiments highlighted a different role for the two MTs in sexual reproduction: the mating events (pairs and clusters) in crosses with sex ratio 10:90 MT-:MT+ were far greater than the mating events in the opposite ratio. This result suggests that the MT+ is the limiting MT regarding the sexual reproduction efficiency. This would be consistent with the time course experiments where the basal expression of *MRP1* in the control MT+ seems to be related to a stronger response of MR genes in the mate perception process.

2.4.5 Gene expression changes over a wider time window

Based on previous studies, I chose genes that I investigated in sexualized and conditioned cells to understand their dynamics and role during the mate perception phase. I explored a new RNAseq dataset, that has recently been generated in the laboratory (R. Annunziata), trying to gain more information by looking at later stages of the process. This exploration did not fill the gap from 6 hours to 24 hours, so it was not possible to confirm a possible delay in the activation of target genes conducible to strain and time variability.

Mating type Related genes

The MR genes in the cross condition were always regulated, and the gene showing the strongest induction was *MRP1*, which was upregulated already at 1 hour, peaked at 24 hours and remained very highly expressed at 5 days (Fig.2.22). This sustained overexpression at 5 days is very interesting, at this stage during reproduction gametes are generally less abundant and parental cells that did not form gametes are arrested in their

cell cycle. *MRP2* in crosses versus MT+ was upregulated and more or less constant in its expression among different time points. For *MRM1* the read counts over time did not change considerably. *MRM2* was also induced in crosses. These data suggest that *MRP1*, *MRP2*, *MRM1* and *MRM2* play a role also beyond the previous investigated time windows (2-6/8 hours). It could be interesting to perturb their function by genetic transformation to understand clearly their role and their mode of interaction.

Cell cycle

CYCB1 was repressed in *S. robusta* MT- cells treated with SIP+, arrested in their cell cycle (Moeys et al. 2016). In this study, cells were followed until 12 hours, and the difference in expression with respect to non-treated cells was observed starting from 6 hours. In *P. multistriata* the filtered medium of MT+ affected the expression of *CYCB1* determining a downregulation at 2 and 6 hours (Table 2.5). An important conclusion from this result is that the primary pheromone present in the conditioned medium might be able to directly block the cell cycle, stopping cells from proceeding to mitosis. *CYCB1* in the cross condition is downregulated as early as 1 hour, confirming this very early signature of growth arrest. Expression however increases at 24 hours, when cells are still arrested, similarly to *CYCB2/2*. This data is less easy to interpret, and it will be important to look at other cyclins and cell-cycle related genes to reconstruct exactly the mechanism mediating the switch between mitosis and meiosis.

Meiotic genes

MRE11 and RAD50 are involved in DNA double-strand breaks repair and in homologous recombination during meiosis (Patil et al. 2015). These genes were found upregulated in *S. robusta* cells treated with SIP+ both between 1 and 2 hours after illumination (Moeys et al. 2016). In this species the combination of the *CYCB1* downregulation with the *MRE11* and *RAD50* upregulations indicated that MT- perceiving SIP+ switched from mitosis to

meiosis. In *P. multistriata MRE11* was induced at 1 hour, similarly to *S. robusta*, while *RAD50* and *RAD51-A1*, another key gene for meiosis, were clearly upregulated at 24 hours. All three genes remained highly expressed at 5 days, indicating that, although most of the gametogenesis occurs in the first 24-48 hours, some of the parental cells at 5 days were still undergoing meiosis.

Signal transduction genes

sGC was strongly upregulated at all time points versus vegetative cells. This expression pattern consolidates the putative important role of *sGC* over the whole process, indicating that the parental cells continue to perceive each other. *GCPDE* was upregulated in comparisons with vegetative MT+, but this reflects likely the different basal expression level in vegetative cells (see histogram of the Figure 18.a). In all three comparisons between crosses it was moderately upregulated, with maximum FC at 5 days versus 1 hour. In *S. robusta* the *GCPDE* gene has a key role in mate perception, indeed treating the MT- with SIP+ its expression increased peaking at 3 h (Moeys et al. 2016). In these RNAseq data, its induction is lower than the induction of sGC. This could mean that in different species the GCs involved in mate perception are different or that in both species both GCs play a role but in different steps of cross talk.

75120, another gene involved in signalling, is induced at 1 hour and remains upregulated, although its levels are never very high, possibly explaining why in my experiments its induction was not always detected (Tables 2.4 and 2.5).

The cathepsin and aardvark genes were strongly regulated in cross comparisons versus both vegetative MTs; the first gene, a pepsin-like aspartate protease, could be involved in mechanisms to ensure the mating efficiency: in *Saccharomyces cerevisiae*, an extracellular peptidase is involved in the degradation of the pheromone creating a density gradient and helping the mating partner to locate the opposite MT (Barkai et al. 1998). It would be

exciting to verify if this happens also in *P. multistriata,* especially considering that MRP1, which is possibly a secreted protein, is a strong candidate for the role of MT+ pheromone. Aardvark is a component of the junctional complex for cell-cell adhesion and it is required for cell signalling (Grimson et al. 2000), therefore it could be involved in adhesion between opposite MTs to ensure gamete fusion.

Nutrient transport

The NarU transporter did not change at 1 hour while it was less expressed at 24 hours and 5 days (Figure 2.18.a). In my experiments this gene was not always downregulated, meaning that NarU is not a good proxy for the early phases of mate perception, since its regulation is most likely a consequence of metabolic changes related to the growth arrest during sexual reproduction. It is likely that, given the strain and time variability, to catch this deregulation it is necessary follow this gene beyond 6-8 hours.

Genes with unknown function

The two non-annotated genes 7488 and 31872 were both strongly upregulated in crosses versus vegetative MTs. In the cross at 1 h 7488 peaked and its upregulation decreased gradually over time, while 31872 peaked at 24 hours, its expression was lower respect to 7488 at each time point.

At the beginning of the cross the opposite MTs start to exchange chemical signals and then start meiosis, while at the same time the cell cycle is arrested in the non-mating cells in the culture. The results presented here also suggest that at 5 days part of population continues the mate perception phase, although this is not evident when inspecting cultures. Sustained mate perception, and the associated changes in the gene expression profile, are possibly explaining why the parental cells do not divide and do not grow for days despite no nutrient limitation is experienced by the cells in co-culture. These RNAseq data support

my choice of 7488 as a molecular marker to test fractions of the medium of MT+ to search for the putative pheromone.

Chapter III:

Sex genes in the environment: exploring TARA Oceans data

3.1 Introduction

Meta omics analysis is the investigation of DNA, RNA, proteins, or metabolites (metagenomics, metatranscriptomics, metaproteomics, metabolomics) of all organisms living in an environment of interest (soil, air, fresh and sea water). Marine studies have advanced enormously in the last two decades thanks to progresses in sequencing technologies. The great interest in marine unicellular eukaryotes is due to many reasons that range from understanding their role in ecosystem functioning, their adaptation to different environmental conditions to their potential use in biotechnology applications. The leap from single genome exploration to metagenomes and metatranscriptomes allows to explore diversity in terms of functions and plankton community composition. In particular, for studies of functional diversity in a plankton community, the metatranscriptome allows to capture all the genes expressed at a given moment by all its members. Meta transcriptomic approaches have been used to address several biological questions. One of the most interesting questions regards the capability of diatoms to live in iron-poor waters. Marchetti et al. (2012) analysed metatranscriptomic data generated from a microcosm experiment in which waters from the north-eastern Pacific Ocean were enriched with iron. The aim was to identify genes regulated in the iron-enriched condition as compared with the iron-limited natural condition. The Authors found that genes encoding rhodopsins were underrepresented under iron-enriched conditions, and that rhodopsin genes were found in the genome of oceanic *Pseudo-nitzschia granii* and *Fragilariopsis cylindrus* but not in the coastal diatoms *Pseudo-nitzschia multiseries*, *Thalassiosira pseudonana*, or *Phaeodactylum tricornutum.* Therefore, they postulated that rhodopsin may play a role in helping diatoms in low-iron condition, allowing the production of ATP from light energy without the involvement of PSI that requires iron to function (Marchetti et al. 2012). In another study, metatranscriptome analyses were used to study pathways of nitrogen and phosphorus metabolism in diatoms in the Narragansett Bay (USA). From this study it emerged that genes responding to the nutrients were different among different diatoms and

during time-series sampling; this could mean that this resource partition prevents competition in the same environment indicating that diatoms modulate their cellular physiology to partition their niche space (Alexander et al. 2015). In these studies, metatranscriptome analyses were aimed to investigate defined biological issues in specific environmental conditions.

NGS and meta 'omics' approaches have paved the way for many appealing and challenging projects among which Tara Oceans (https://oceans.taraexpeditions.org/en/). The Tara expedition represents the first attempt to provide molecular resources to explore biological questions at a global spatial scale, covering the whole planktonic community over a broad range of environmental conditions (Pesant et al. 2015). The Tara oceans expedition involved a large team of researchers and run for four years, from 2009 to 2013, collecting over 35,000 samples in 210 different stations (Figure 3.1) in the world oceans (Pesant et al. 2015). Besides information on the planktonic organisms present at the different sampling stations, data on genetic diversity, metatranscriptomic and metagenomics data were collected together with a set of environmental parameters.

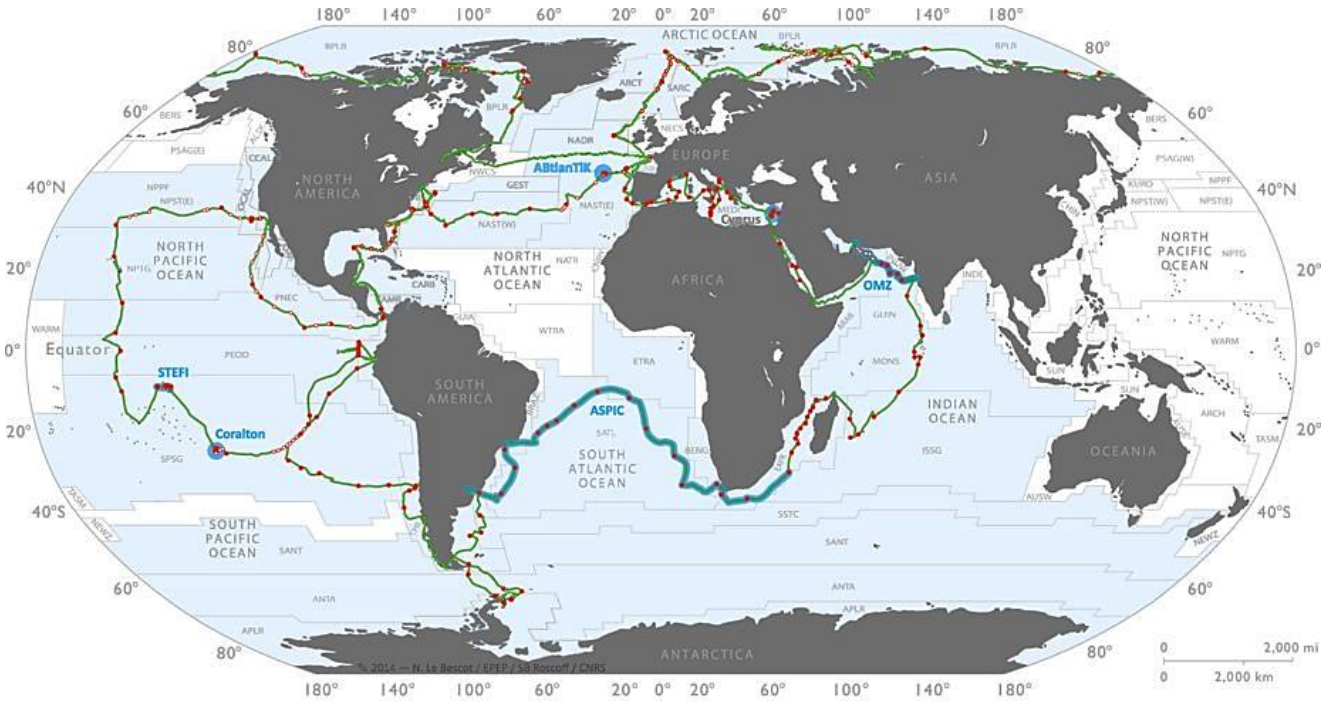


Fig.3.1. Sampling route of the Tara Oceans Expedition (green track), showing stations where plankton were sampled in their environmental context (full red dots) and where only environmental parameters were measured (open red dots). Topical experiments are identified along the sampling route (light blue). Longhurst biogeographical provinces are shown in the background and those sampled during Tara Oceans Expedition are highlighted in blue (from Pesant et al. 2015).

The baseline was to carry out regular sampling but, in addition, specific experiments (topical experiments) were performed to study specific processes in the oceans (Figure 3.1). For example, a topical experiment was conducted to study longitudinal transport by Agulhas rings across the South Atlantic Ocean (Villar et al. 2015).

Samples were collected at different depths:

- The surface water layer (SRF, SUR, SURF), that is the layer between 3 and 7 m below the sea surface.

- The deep chlorophyll maximum layer (DCM), that is the subsurface layer in the water column where chlorophyll concentration reaches its maximum.

- The mesopelagic zone (MESO) that is the layer between 200 and 1000 m depth.
- The oxygen minimum zone (OMZ) that is the layer where the oxygen saturation is lowest (sampled only in selected sites).
- The epipelagic mixing layer (MIX), that is the upper part of the marine water column with generally uniform properties.

The organisms were sampled separating them in different size fractions:

- < 3 µm for viral and bacterial fraction
- 0.8 - 5 µm for pico and small nanoplankton
- 5 - 20 µm for nanoplankton
- 20 – 180 µm for microplankton
- 180 - 2000 µm for mesoplankton

This huge amount of samples represents an important resource for discovering new taxa, non-annotated genes and to obtain a complete assessment of marine biodiversity.

The biodiversity of eukaryotic plankton in the photic zone was explored using the hypervariable V9 region of 18S rDNA (de Vargas et al. 2015) at 47 stations over the whole size range of planktonic organisms. The V9 metabarcodes were clustered to obtain biologically meaningful operational taxonomic units (OTUs). The obtained OTUs were about 150,000 versus the 11,000 OTUs formally described; one third of the total OTUs were not assigned to known eukaryotic groups. Furthermore, the eukaryotic taxonomic diversity was higher in the smaller fractions. About 87,000 assigned OTUs were classified into 97 taxa groups, some of these OTUs were assigned to planktonic eukaryotic species (about 11,200 species). Comparing these data, it emerged that for most part of taxa the number of OTUs was three to eight times greater than the number of species known from morphologically-based taxonomy. This abundance of OTUs with respect to known taxa

could be explained by the existence of cryptic species. As for diatoms, the group for which the highest number of reference sequences is available (1,232 sequences), many new rDNA sequences were discovered. These new rDNA sequences grouped both in clades for which reference sequences were available and in new clades.

The number of diatom species was estimated to range from a minimum of 30,000 to a maximum of 100,000 by Mann and Vanormelingen (2013). Undoubtedly, the TARA data set represents an opportunity to better understand the true diversity number of specific taxa. Malviya et al. (2016) have exploited the potential of the TARA data set to highlight the diatom diversity and its global distribution in the oceans, using V9-18S rDNA sequences from 293 fractions of 46 stations. The ribotype abundance showed that the diatoms were the second eukaryotic lineage, the first being the Dinophyceae. Diatom reads were about 5% of protist ribotypes and about 3% of total eukaryotic reads. In some stations there was a large abundance of diatoms, for example in the TARA_011 station in the Mediterranean Sea where their contribution reached about 75% in the photosynthetic community, and in the TARA_84 station (Southern Ocean), where it was more than 78%. On the contrary, some stations like TARA_18, TARA_20 and TARA_30 (Mediterranean Sea) had very low percentages of diatoms with respect to other photosynthetic taxa. Within diatoms 79 genera were found, of which *Chaetoceros* was the most abundant genus with about 23% of assigned sequences, followed by *Fragilariopsis* (15.5%), *Thalassiosira* (about 14%), *Corethron* (11%), *Leptocylindrus* (about 10%), *Actinocyclus* (about 9%), *Pseudo-nitzschia* (4.4%) and *Proboscia* (about 4%). Among these genera the diversity was very variable, spanning from the lower diversity in *Asteroplanus* (one ribotype) to the highest diversity in *Chaetoceros* (6,904 ribotypes). The greatest number of ribotypes was found in fractions with size from 5 to 180 μm, but a conspicuous number was found also in smaller size fractions. Ribotypes attributed to larger species were found in the smaller size fractions, probably on account of broken cells, faecal pellets, or gametes. The presence of

ribotypes in the 180- 2000 μm size fraction was most likely due to the presence of diatom chains but also to small cells ingested by larger organisms or associated with them or with microplastics. *Pseudo-nitzschia*, with more than 103 ribotypes, was mostly found in the 5-20 μm size fraction and with a lesser number of ribotypes in the 0.8-5 and 20-180 μm size fractions. Few *Pseudo-nitzschia spp.* were present in the 180-2000 μm size fraction. Regarding depth, this genus was almost equally present in surface and DCM samples. *Fragilariopsis* was mainly abundant in the fraction with size from 0.8 to 20 μm and mainly found in SUR depth. Unassigned Bacillariophyta, with about 104 OTUs, was found mostly in the 5-20 μm size fraction and about 75% in SUR depth. Furthermore, a comparison between light microscopy observation and V9 ribotype counts was performed for 15 stations (Malviya et al. 2016). In these stations (except 072 and 085) V9 ribotypes assigned to *Pseudo-nitzschia* genus were found, and these results were confirmed by light microscopy observations. Only for station 084 the presence of *Pseudo-nitzschia* was not confirmed in light microscopy, but TARA_084 presents inconsistencies also for other comparisons. Indeed, light microscopy observation recorded about 90% of *Fragilariopsis* while V9 ribotypes were assigned in equal percentage to *Fragilariopsis*, *Chaetoceros* and other Centric.

In terms of gene discovery, Tara was also an opportunity to create a new and exhaustive gene catalog for marine organisms that is available in two databases, the Ocean Microbial Reference Gene Catalog (OM-RGC) and the Marine Atlas of Tara Oceans Unigenes (MATOU), which only recently became public at http://tara-oceans.mio.osupytheas.fr/ocean-gene-atlas/ (Ocean Gene Atlas, OGA).

The OM-RGC is a merged database (Sunagawa et al. 2015) where metagenomes of organisms up to 3 μm (viruses, prokaryotes, and picoeukaryotes), from 243 samples from 68 stations sampled within Tara, were combined with public data sets of genes. These datasets are:

- Global Ocean Sampling (GOS) expedition (Williamson et al. 2008)
- Pacific Ocean Virome (POV) project (Hurwitz and Sullivan 2013)
- NCBI reference genomes (relevant to the marine environment)
- Moore Microbial Genome Sequencing Project: phage/viral genomes (MPVG)

In OM-RGC there are more than 40 million non-redundant genes and a large percentage (81.4%) of these genes came from Tara Oceans samples, about 28% of the genes were not taxonomically annotated and were associated, mainly, to viral fractions (< 0,22 µm). The percentage of non-annotated genes decreases in prokaryote and eukaryote enriched samples (Sunagawa et al. 2015).

The same effort has been done also for organisms sampled in fractions with size from 0.8 up to 2000 µm present in the surface and DCM, generating the Marine Atlas of Tara Oceans Unigenes (MATOU) (http://www.genoscope.cns.fr/tara/; Tara Oceans Coordinators et al. 2018). In this study the samples were represented by 441 Tara Oceans samples. In this catalog there are 116 million unigenes of which about half with taxonomic assignation. The association between unigenes and taxa was possible thanks to the reference database built assembling UniRef90 (ftp://ftp.uniprot.org/pub/databases/uniprot/uniref), the MMETSP dataset (Marine Microbial Eukaryote Transcriptome Sequencing Project) and Tara Oceans Single-cell Amplified Genomes. To these databases were added three Rhizaria transcriptomes (available in the European Nucleotide Archive) and transcriptomes of the epipelagic copepod *Oithona nana* (Madoui et al. 2017). The authors employed MATOU to give an example of how the annotated gene catalog can be useful for studying environmental gene expression (Tara Oceans Coordinators et al. 2018). In this study the analysis of the Pfam domains of the ferredoxin (PF00111) and flavodoxin (PF00258) genes was performed, looking at relative levels in the Chlorophyta, Pelagophyceae, Haptophyceae,

Bacillariophyta and Dinophyceae by calculating the ratio of their abundances and expression of target genes. In all groups, except diatoms, there were few variations in gene abundances and weak correlations with iron concentrations, while the ratios of relative expression showed strong variations, particularly for Chlorophyta, Pelagophyceae and Haptophyceae, indicating that these three groups modulate the relative levels of ferredoxin and flavodoxin mRNAs. In diatoms, the level of flavodoxin mRNA was more than ferredoxin mRNA, even if in the few coastal stations, the diatoms showed a strong upregulation of ferredoxin genes. Diatoms displayed in metagenome more fluctuations in the ferredoxin/flavodoxin content, suggesting that single diatom species are adapted to iron regimes in different specific ways, contrary to what was observed in haptophytes, chlorophytes and pelagophytes. These results suggest that photosynthetic taxa are adapted to nutrient limitations in different ways, at genomic and/or transcriptional levels. So, MATOU could be a resource to study the molecular responses of plankton to environmental changes, investigating transcript regulation or gene copy numbers implicated in the responses (Tara Oceans Coordinators et al. 2018). In this chapter, I exploited the MATOU database to look at the distribution and abundances of seven genes related to sexual reproduction and meiosis. I considered the five mating type related (MR) genes of *P. multistriata* that are conserved in *Pseudo-nitzschia* and *Fragilariopsis* species (Russo et al. 2018) (Figure 3.2; see Chapter 1 for an illustration of these genes). I tested if MR genes are present also in other diatom species or genera.

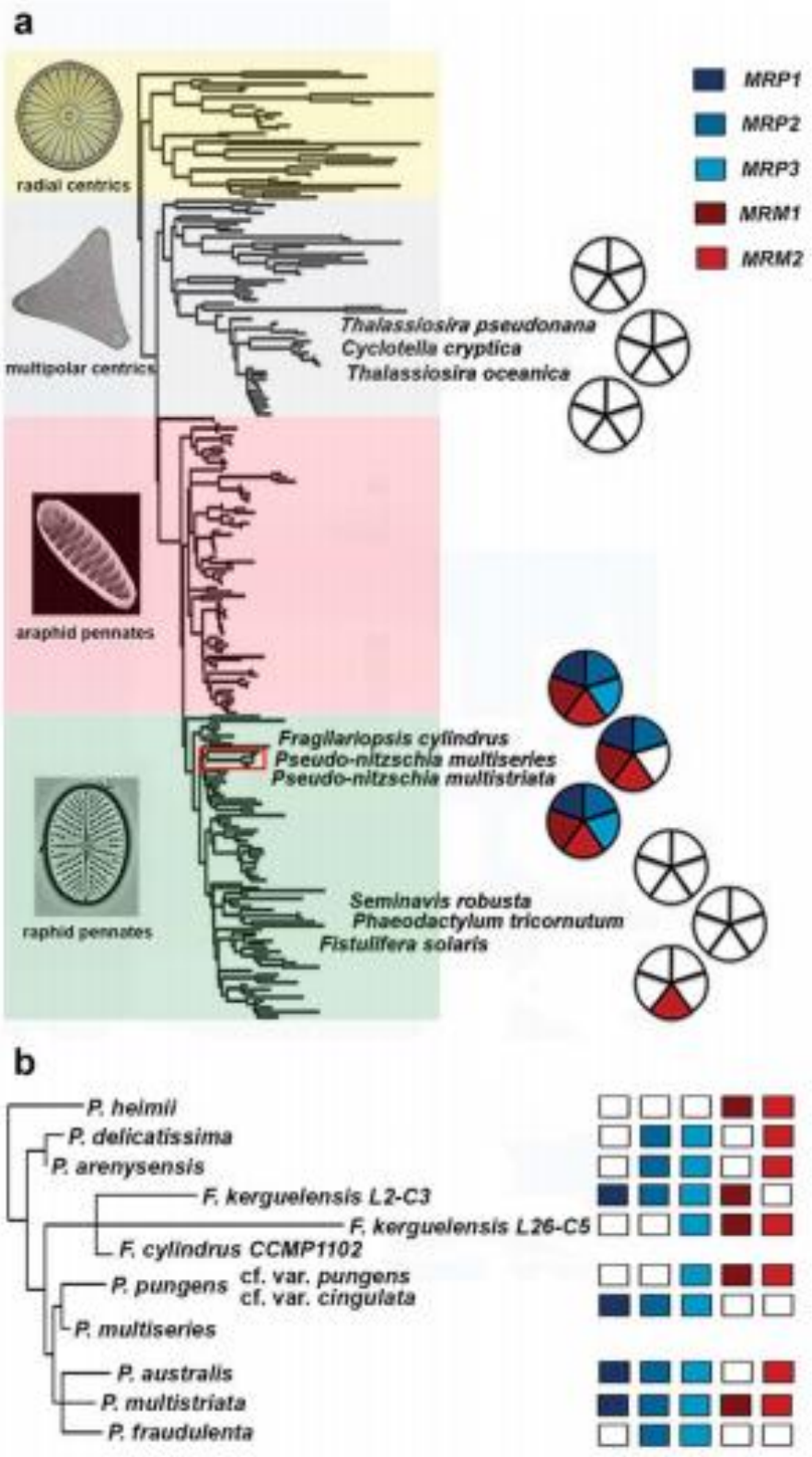


Fig.3.2 *P. multistriata* MT-biased genes conservation. **a** Conservation in diatom genomes. The diatom species with a sequenced genome are shown in a simplified diatom phylogenetic tree based on 18S, colour in the pie next to each species indicates the presence of a homolog, white indicates that no homology had been detected. Shades on the tree demarcate diatom lineages indicated on the left. The red rectangle indicates the area that is magnified in **b**. **b** Conservation in the transcriptomes of *Pseudo-nitzschia* and *Fragilariopsis* species. From Russo et al. (2018).

A conserved meiotic toolkit was identified for species that have sexual reproduction (Ramesh et al. 2005) and it could represent an important resource to find universal marker genes to detect sex events. For diatoms, Patil et al. (2015) explored five diatom genomes, those of *Thalassiosira pseudonana, Phaeodactylum tricornutum, Fragilariopsis cylindrus, Pseudo-nitzschia multiseries, Pseudo-nitzschia multistriata* and the *de novo* transcriptome of *Seminavis robusta* to identify the conserved meiotic toolkit in common among these species. In this study homology searches for 60 meiotic proteins were performed. Within this dataset, there were both genes that are known to be specific of the meiosis process and genes that are involved in meiosis but that are expressed also during mitosis. A total of 42 out of 60 genes were found in all examined diatoms, and 15 of them are known to be exclusive of meiosis. Only five of them were present in all the six investigated species. These genes were: *SPO11-2, MND1, MSH4, MSH5* and *MER3*. RNAseq data relative to *S. robusta* strains of opposite MTs were collected at several time points during and after meiosis in co-cultures of the two parental strains undergoing sexual reproduction and in the two MTs in mono-culture. From this RNAseq, cpm (read counts in counts per million) values of transcripts relative to the 42 meiotic genes were displayed in a heatmap showing clearly that these genes were highly expressed during meiosis, with a few exceptions (*SPO11-3/Top VIA, XRCC, MCM6* and *MCM7*) (Patil et al. 2015). In *P. multistriata* the gene expression of selected meiosis genes, *RAD21, SPO11-2* and *RAD51* paralogs, was investigated by qPCR: *RAD21, SPO11-2, RAD51-A*, and *RAD51-C* were upregulated during meiosis and, among these, *SPO11-2* and *RAD51-A1* were the most highly expressed during sexual reproduction (Patil et al. 2015). *SPO11-2* is a gene involved in double strand break formation in homologous chromosomes in meiotic recombination of plants, while *RAD51-A1* is involved in homologous pairing, strand invasion, and in DNA repair mechanisms (Patil et al. 2015). These two genes are most likely required during meiosis in all diatom species, and they could be good candidate markers for events of sexual reproduction. Here, I explored the potential of the TARA data taking advantage of our

knowledge about sexual reproduction in *Pseudo-nitzschia multistriata*, focusing on the five MR genes and the two markers of meiosis, *SPO11-2* and *RAD51-A1*.

These analyses are aimed at:

i) Searching for possible homologs in other diatom species

ii) Exploring the spatial distribution of these homologs in the ocean

iii) Exploring the potential role of the MR genes and the selected meiotic genes like molecular markers to detect sexual reproduction events in phytoplankton.

3.2 Materials and Methods

3.2.1 BLAST analyses and abundances

The predicted proteins of the *MRP1*, *MRP2*, *MRP3*, *MRM1*, *MRM2*, *SPO11-2* and *RAD51-A* (Table 3.1) genes were the queries in blastp searches available on the Ocean Gene Atlas (OGA; http://tara-oceans.mio.osupytheas.fr/ocean-gene-atlas/) (Tara Oceans Coordinators et al. 2018). The expected e-value threshold was imposed equal to $1E^{-10}$, so all the sequences above this value were not included in the results. The OGA website allows to explore also the distribution and abundance of target sequences among all sampled stations. The abundance was fixed like “percent of total genes per sample”, that is gene read coverage in RPKM (Reads Per Kilobase covered per Million of mapped reads) divided by the sum of the total gene coverage for the sample. So the abundance results are the fractions of homologs of all genes in the whole sample.

Tab.3.1 The list of protein sequences searched in the TARA database.

Source	Gene Name	Gene ID	Gene Function
Russo et al, 2018	*MRP1*	PSNMU-V1.4_AUG-EV-PASAV3_0024820	Unknown
Russo et al, 2018	*MRP2*	PSNMU-V1.4_AUG-EV-PASAV3_0122240	Leucine-rich repeat (LRR) trasmembrane region
Russo et al, 2018	*MRP3*	PSNMU-V1.4_AUG-EV-PASAV3_0020770	Unknown
Russo et al, 2018	*MRM1*	PSNMU-V1.4_AUG-EV-PASAV3_0041130 PSNMU-V1.4_AUG-EV-PASAV3_0085380	Heat shock factor protein
Russo et al, 2018	*MRM2*	PSNMU-V1.4_AUG-EV-PASAV3_0006960	Leucine-rich repeat (LRR) trasmembrane region
Patil et al, 2015	*SPO11-2*	PSNMU-V1.4_AUG-EV-PASAV3_0108120	Creates double-strand breaks in homologous chromosomes in meiotic recombination
Patil et al, 2015	*RAD51-A1*	PSNMU-V1.4_AUG-EV-PASAV3_0056780	Double-strand break repair (recombinational repair)

3.2.2 Trees

The retrieved sequences were aligned using MUSCLE (Edgar 2004) with default parameters, the alignment was edited manually removing sequences to have at least one amino acid in common among all sequences. In the analysis sequences from the Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP) data set (Keeling et al. 2014), joint genome institute (JGI) and National Center for Biotechnology Information (NCBI) were included to have a better resolution of the relationship between the query and sequence results. The final alignment was used for phylogenetic analysis with MEGA 7 (Kumar et al. 2016), to infer evolutionary relationships among sequences. The trees were inferred using the Neighbor-Joining method (Saitou and Nei 1987) setting 10000 replicates for the bootstrap tests. The evolutionary distances were computed using the JTT matrix-based method and are in the units of the number of amino acid substitutions per site. Final Trees were edited using in FigTree, the editing was relative to the layout of trees and gene

IDs of sequences that were manually renamed with their taxonomic assignation. The clades with lower bootstraps were collapsed.

3.2.3 Maps

The maps were built by R studio using ggplot2 and ggrepel library packages. The geographical coordinates of TARA stations are available on the PANGEA site https://doi.pangaea.de/10.1594/PANGAEA.842237. The data relative to the distribution of genes were retrieved from OGA.

3.3 Results

3.3.1 Phylogenetic analyses of MR, SPO11-2 and RAD51-A proteins

3.3.1.1 MRM1

From the blast analysis 51 hits in TARA stations were found and, after applying filters described in the Methods, only 25 sequences were used for phylogenetic analyses. The tree (Figure 3.3) shows two main clades, one with all sequences assigned to the order Pelagomonadales and the other clade mostly with sequences assigned to the class Bacillariophyceae. This latter clade contains the query sequence of *MRM1* and the other sequences available in public databases (see Materials and Methods). Based on the tree topology, there are six sequences that are most likely *Fragilariopsis* transcripts (grey box in Figure 3.3), while one transcript annotated as Bacillariales could be either a *Fragilariopsis* or a *Pseudo-nitzschia* transcript.

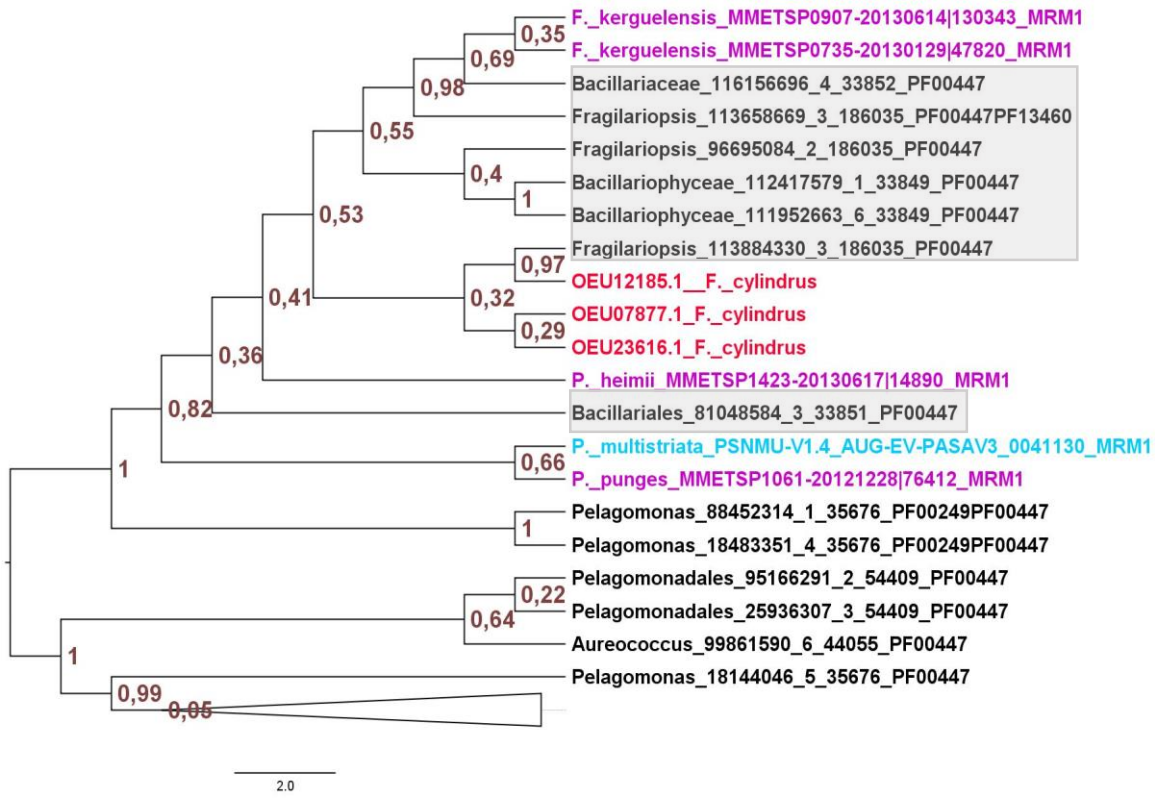


Fig.3.3 Molecular phylogenetic analysis with the Neighbor-Joining method for the MRM1 protein. In black sequences retrieved from the Tara database, in pink from MMETSP, in red sequences from NCBI and in light blue the MRM1 sequence of *P. multistriata.* The grey shadow highlights the TARA transcripts considered to be homologs of the query sequence. The triangle is a collapsed clade.

3.3.1.2 MRM2

The blast analysis retrieved 7403 hits in the TARA database. The phylogenetic analysis was performed with 243 sequences from TARA, six homologous sequences from the MMETSP dataset and three sequences from NCBI. The query sequence MRM2 falls into the clade with three sequences from MMETSP (*P. pungens, P. delicatissima* and *P. heimii*) and with three sequences from the TARA blast results. Two of these sequences were classified as belonging to the order Bacillariales and grouped with known sequences of *Pseudo-nitzschia sp,* while the third was annotated as an *Asterionellopsis glacialis* sequence (Figure 3.4). Bootstrap values in other parts of the tree were too low to allow unambiguous identification of homologues.

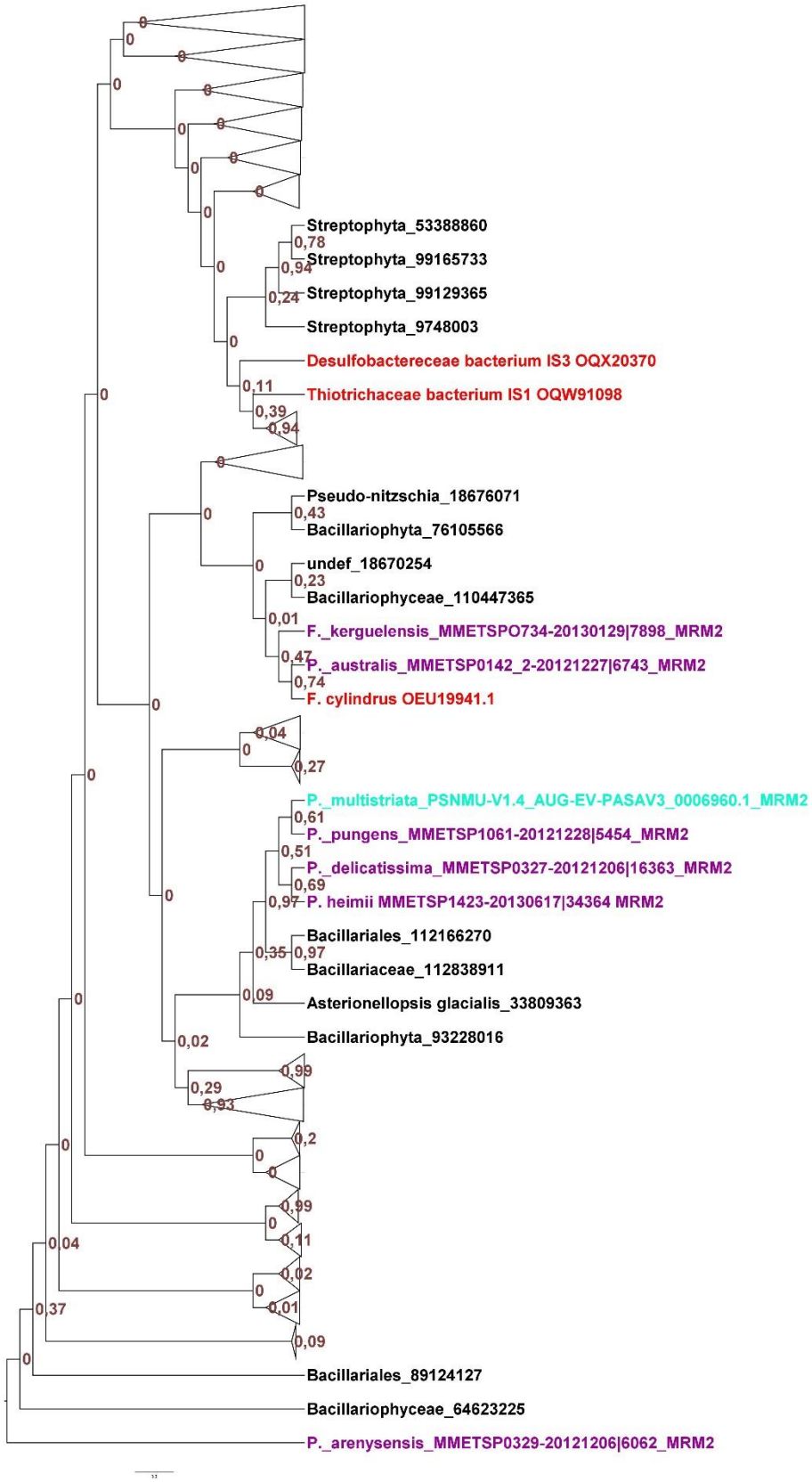


Fig.3.4 Molecular phylogenetic analysis with Neighbor-Joining method for the MRM2 protein. In black sequences retrieved from the Tara database, in pink MRM2 sequences from MMETSP, in red sequences from NCBI and in light blue the MRM2 sequence of *P. multistriata*. The grey shadow highlights the TARA transcripts considered to be homologs of the query sequence. The triangles are collapsed clades.

3.3.1.3 MRP1

The blast analysis retrieved 54 hits in the TARA database that were aligned. 20 out of 54 were used for phylogenetic analysis, these sequences were assigned to the class Bacillariophyceae (Figure 3.5), none of these sequences was assigned at the genus level. Although the bootstrap values are low at some branches, all sequences falling in the clade containing MRP1 sequences identified in Russo et al., 2018, were retained for further analyses (grey box in Figure 5) since MRP1 is known to evolve rapidly (Basu et al. 2017).

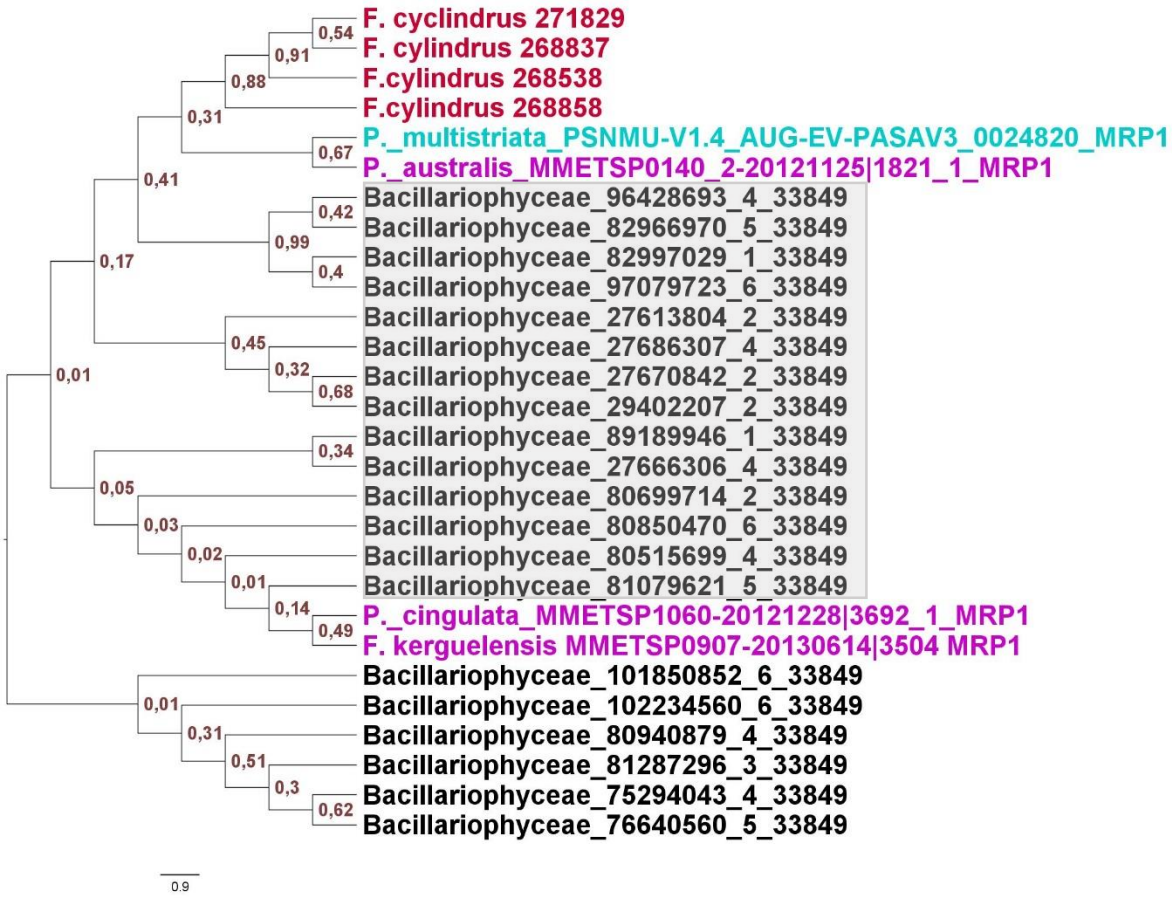


Fig.3.5 Molecular phylogenetic analysis with Neighbor-Joining method for the MRP1 protein. In black sequences retrieved from the TARA database, in pink MRP1 sequences from MMETSP, in red sequences from NCBI and in light blue the MRP1 sequence of *P. multistriata*. The grey shadow highlights the Tara transcripts considered to be homologs of the query sequence.

3.3.1.4 MRP2

The blast analysis gave 3196 hoursits in the TARA database and only 23 were used for phylogenetic analyses. In the tree (Figure 3.6) one well supported clade contains TARA sequences annotated as *Thalassiosiraceae* and *Thalassiosira sp*. A second clade contains *P. multistriata* MRP2, five sequences from the MMETSP database and four sequences from TARA, all likely belonging either to *Fragilariopsis or Pseudo-nitzschia* (grey boxes in Figure 3.6). Sequences assigned to Coscinodiscophyceae, *Asterionellopsis glacialis*, Streptophyta and Bacillariophyta were also retrieved and formed a separate clade with low bootstrap values.

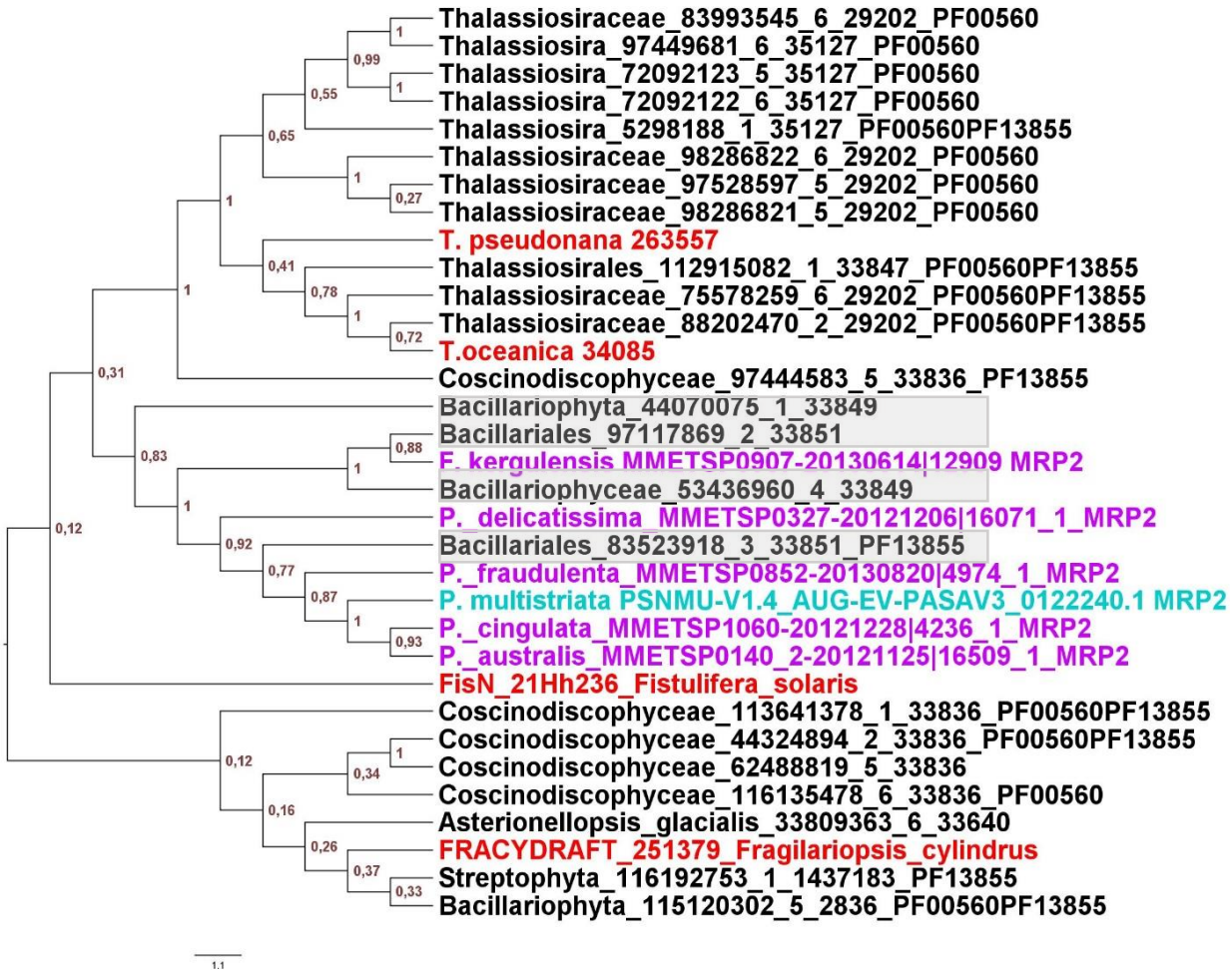


Fig.3.6 Molecular phylogenetic analyses with Neighbor-Joining method for the MRP2 protein. In black sequences retrieved from the TARA database, in pink MRP2 sequences from MMETSP, in red sequences from NCBI and in light blue the MRP2 sequence of *P. multistriata*. The grey shadow highlights the TARA transcripts considered to be homologs of the query sequence.

3.3.1.5 MRP3

From the blast search only 7 hits were found in the TARA database, two of which assigned at genus level, as *Fragilariopsis* and *Pseudo-nitzschia spp*. With such a low number of sequences a phylogenetic analysis would be little informative. In this case I performed an alignment (Figure 3.7), with 6 TARA sequences, excluding the 7th hit as it was too short and too divergent, and including also 6 sequences from the MMETSP dataset. In the alignment there are 24 conserved sites of which 20 present in *Pseudo-nitzschia sp* sequences; eleven of these sites were close in a pattern. Furthermore, there were 9 sites with amino acids with similar chemical properties.

Fig.3.7 Alignment of MRP3 homologs. The 24 conserved amino acids are highlighted in yellow, while the nine similar amino acids are highlighted with a black frame.

3.3.1.6 SPO11-2

A total of 250 hits were found when the TARA database was queried with *SPO11-2* and 111 sequences were used for phylogenetic analyses. In the phylogenetic tree (Figure 3.8) the *SPO11-3* and *SPO11-2* genes of *Thalassiosira pseudonana*, *Phaeodactylum tricornutum*, *Fragilariopsis cylindrus*, *Pseudo-nitzschia multiseries*, *Pseudo-nitzschia multistriata* and *Seminavis robusta* were included (Patil et al. 2015). The presence of *SPO11-3* ensures a better resolution of the tree. Many sequences from the TARA database grouped in a big clade separated from few TARA sequences that grouped with *SPO11-2* sequences. This latter clade contains 5 sequences from TARA, 4 assigned to the Bacillariophyta phylum and 1 annotated as Thalassiosiraceae.

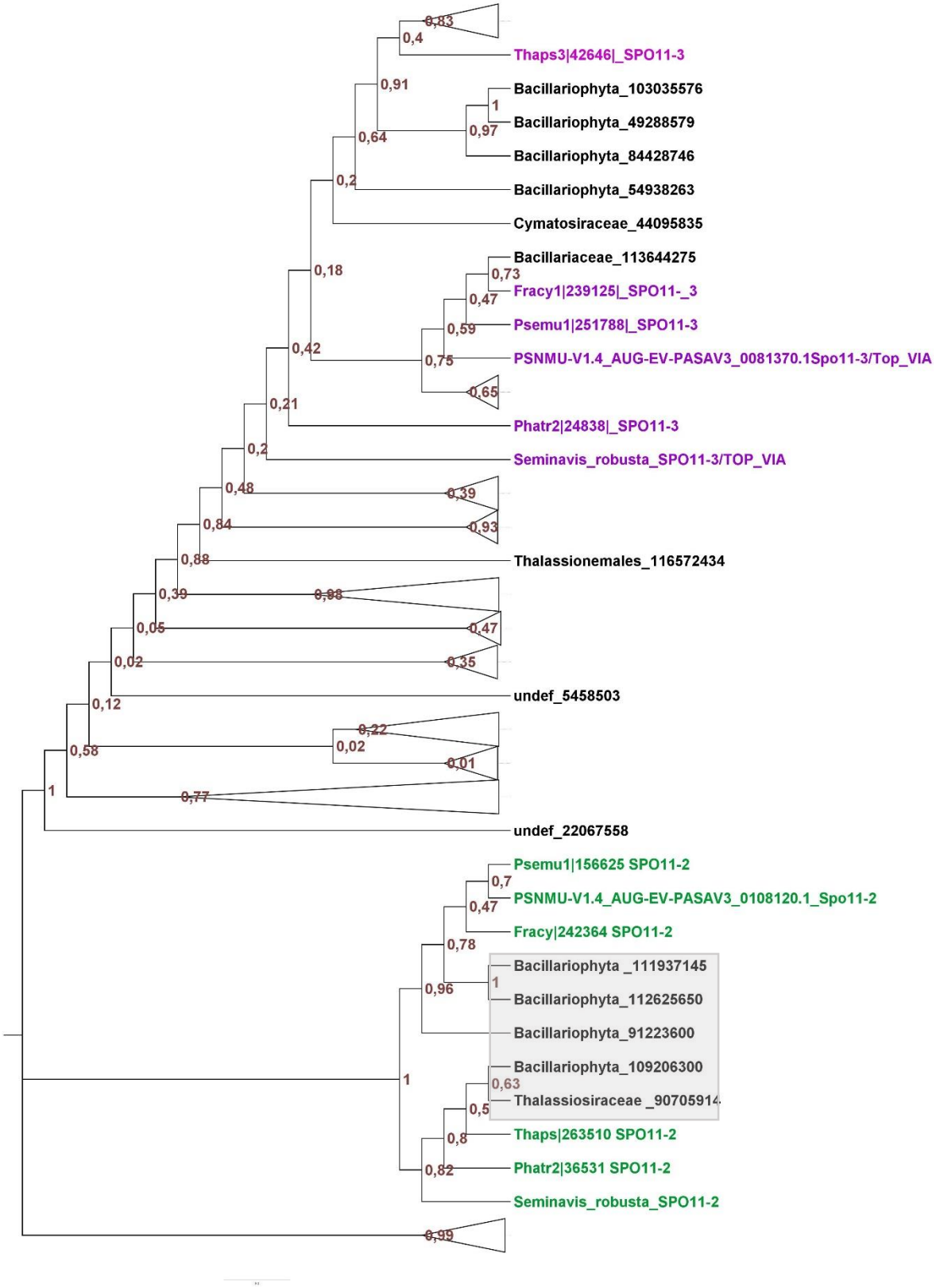


Fig.3.8. Molecular phylogenetic analysis with Neighbor-Joining method for the SPO11-2 protein. In black sequences retrieved from the TARA database, in pink the SPO11-3 sequences and in green the SPO11-2 sequences. The grey shadow highlights the TARA transcripts considered to be homologs of the query sequence. The triangles are collapsed clades.

3.3.1.7 RAD51-A1

When *RAD51-A1* was used as query, 4554 hits were found in the TARA database. Many sequences with taxonomic path including the term “Metazoa” (for example: Biota; Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Crustacea; Maxillopoda; Copepoda; Calanoida; Calanidae) were removed to have a lower number of sequences. The alignment was edited manually obtaining a file with 135 sequences, including sequences of *RAD51-A1*, *RAD51-A2*, *RAD51-B* and *RAD51-C* belonging to other diatom species, *Ectocarpus siliculosus* and *Homo sapiens* (Patil et al. 2015).

In the tree (Figure 3.9) *RAD51-A1/2*, *RAD51-B* and *RAD51-C* grouped in three different clades. TARA hits grouped mainly with *RAD51-A1/2* sequences but the bootstrap values were very low.

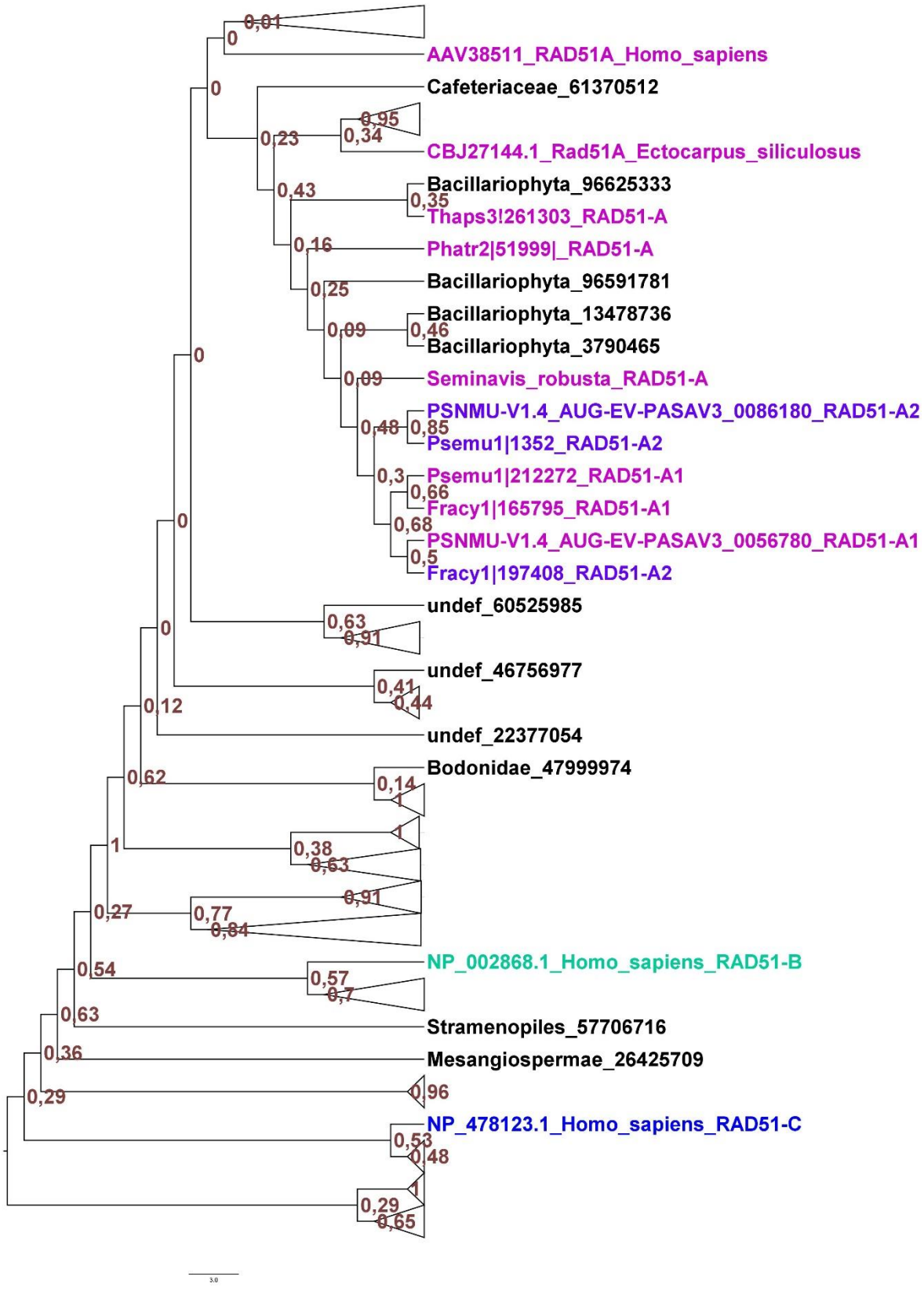


Fig.3.9. Molecular phylogenetic analysis with Neighbor-Joining method for the RAD51 proteins. In black sequences retrieved from the TARA database, in pink the RAD51-A sequences, in purple RAD51-A2 sequences, in green RAD51-B sequences and in blue RAD51-C sequences. The triangles are collapsed clades.

3.3.2 Abundances of putative homologs

From now on, all results shown are relative to genes considered to be homologs to the queries based on the results of the phylogenetic analyses, shaded in grey in the trees in Figures 3.3-3.6 and 3.8.

Tables with abundance of genes in each station where they were recovered are shown in the Appendix 1. The sample with the highest abundance (0.0004%) of *MRM1* sequences was N000001440 (5-20 µm fraction) that belongs to station TARA_084 (Southern Ocean). The major contribution to this percentage comes from *Fragilariopsis sp* (113658669 gene ID). For *MRM2*, the sample with the highest abundance was N000001440 (5-20 µm fraction) of the same TARA_084 station, with percentage abundance of about 0.00015%. The major contribution to this percentage comes from gene 112838911 (*Bacillariaceae* family) that could belong to the genus *Pseudo-nitzschia* (Figure 3.4). For *MRP1* the sample with the highest abundance was A200000123 (0.8-5 µm fraction) in the station TARA_007 (Mediterranean Sea) with percentage abundance of about 0.0012%. The following samples were N000001823 and N000001656, belonging, respectively, to TARA_111 (20-180 µm fraction, South Pacific Ocean) and TARA_102 (5-20 µm fraction, South Pacific Ocean). Among all genes, annotated as Bacillariophyceae, the genes with higher abundances (about 0.0003%) were two (27666306 and 27686307). For *MRP2* and *MRP3* the samples with the highest abundance were N000001438 and N000001440. The first belongs to the 0.8-5 µm fraction and the second to the 5-20 µm fraction; both samples are from TARA_084. The abundance for *MRP2* was relative to one gene (53436960) assigned to the class Bacillariophyceae that could be a *Fragiliaropsis* transcript (Figure 3.6). The sample with the highest abundance of *MRP3* was N000001440 of the station TARA_084, 5-20 µm fraction. The abundance of *MRP3* was almost totally due to the presence of gene 113648004 annotated as *Fragiliaropsis sp*.

The sample with the highest abundance of *SPO11-2* was N000000531 that belongs to TARA_064 station (5-20 µm fraction, Indian Ocean). The most abundant homologue of *SPO11-2* was gene 90705914 annotated as Thalassiosiraceae. The bootstraps in the tree for of *RAD51-A* homologs were very low (Fig. 3.9), so the abundances were not considered.

In Figure 3.10 the total percentage abundance of all homologs contained in all samples of each station, for each MR gene are shown. The details of homolog abundances for each MR gene and for each station are shown in the tables reported in the Appendix 1.

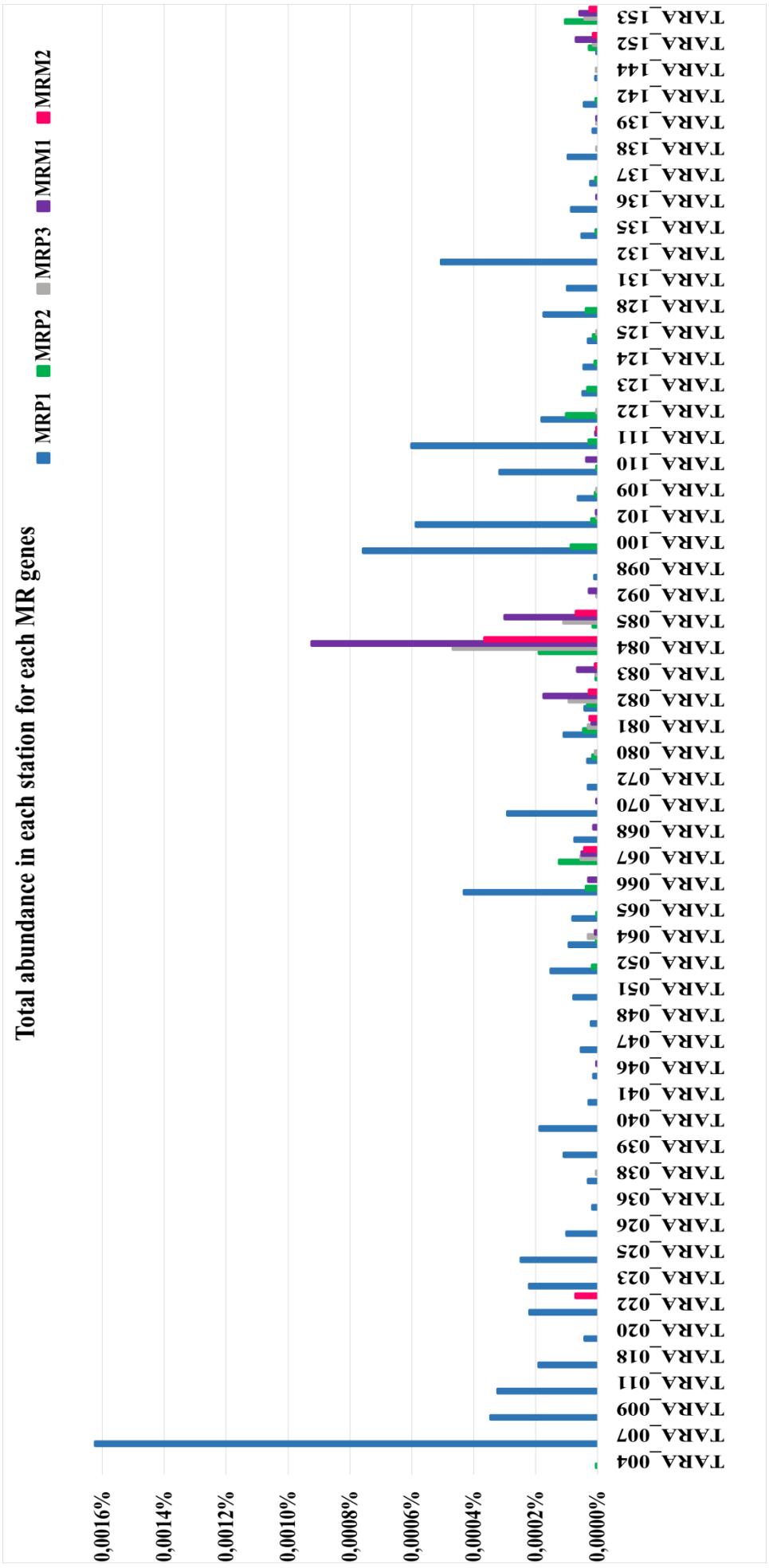


Fig.3.10 Total abundances of MR homologs in the TARA stations.

3.3.3 Global distribution of transcripts

Homologs of *MRM1* were recorded in 19 stations, while those of *MRM2* in 10 stations (Table 3.2). Homologues of *MRP1*, *MRP2* and *MRP3* were recorded in 50, 27 and 19 stations, respectively. Comparing the stations where MR genes were found, there were only four stations in which all MR genes were detected (Table 3.2): the stations TARA_081, TARA_067, TARA_152 and TARA_082. There were stations where only homologues of a single gene were recorded, in particular, there are 19 stations where only *MRP1* was detected and one where only *MRP2* was detected.

Tab.3.2. Tara stations where different combinations of MR genes were recorded.

Combinations of MR genes	**n° of stations**	**Station IDs**
MRM1 MRM2 MRP1 MRP2 MRP3	4	TARA_081 TARA_067 TARA_152 TARA_082
MRM1 MRM2 MRP1 MRP2	1	TARA_111
MRM1 MRM2 MRP2 MRP3	4	TARA_083 TARA_153 TARA_084 TARA_085
MRM1 MRP1 MRP2 MRP3	2	TARA_064 TARA_102
MRM1 MRP1 MRP2	2	TARA_066 TARA_110
MRM1 MRP1 MRP3	1	TARA_139
MRP1 MRP2 MRP3	4	TARA_125 TARA_080 TARA_109 TARA_122
MRM1 MRP1	4	TARA_070 TARA_068 TARA_046 TARA_136
MRM1 MRP3	1	TARA_092
MRM2 MRP1	1	TARA_022
MRP1 MRP2	9	TARA_128 TARA_137 TARA_100 TARA_135 TARA_052 TARA_065 TARA_124 TARA_142 TARA_123
MRP1 MRP3	3	TARA_038 TARA_138 TARA_144
MRP1	19	TARA_132 TARA_131 TARA_007 TARA_039 TARA_072 TARA_011 TARA_098 TARA_023 TARA_026 TARA_048 TARA_009 TARA_025 TARA_040 TARA_020 TARA_047 TARA_041 TARA_051 TARA_018 TARA_036
MRP2	1	TARA_004

The distribution pattern for the 5 *MR* genes is shown on a map (Figure 3.11), where for each station there is a pie chart with colours for each gene; in the Appendix 2 single maps for each putative MR homolog are shown.

The *SPO11-2* gene was found in 10 stations (Table 3.3 and Figure 3.11), and in all these stations MR genes were also present.

Tab.3.3 The stations where *SPO11-2* was detected, in the table are reported also the recovered MR genes.

Stations	SPO11-2	MRP1	MRP2	MRP3	MRM1	MRM2
T_064	X	X	X	X	X	
T_067	X	X	X	X	X	X
T_081	X	X	X	X	X	X
T_082	X	X	X	X	X	X
T_083	X		X	X	X	X
T_084	X		X	X	X	X
T_085	X		X	X	X	X
T_142	X	X	X			
T_152	X		X	X	X	X
T_153	X	X	X	X	X	X

In table 3.4 are shown all stations for which DNA metabarcoding data indicate the presence of *Pseudo-nitzschia* and/or *Fragilariopsis* species, and for each station the presence/absence of MR genes is reported. I looked for the presence/absence of *Pseudo-nitzschia* and *Fragilariopsis* exploring single "Krona interactive pie charts" for all stations examined by (de Vargas et al. 2015) (http://taraoceans.sb-roscoff.fr/EukDiv/). Krona organizes metagenomics data for hierarchical visualization based on abundance of sequences, with multi-layered pie charts explorable by zooming. I integrated the data from de Vargas et al. (2015) with: i) data reported in (Malviya et al. 2016), where the presence of species detected through DNA metabarcoding was confirmed by light microscopy (LM) observation and with ii) additional data on species identification in LM provided by

Eleonora Scalco (SZN), involved in the TARA project. For the stations 83, 142, 152 and 153, listed in Table 3.3, no information on species composition was available.

In Figure 3.12 there are two graphs where stations containing an MR gene are distributed according to latitude and temperature. *MRP1*, *MRP2*, *MRP3* and *MRM1* show more or less the same distribution for latitude, while *MRM2* seems to be absent in latitudes ranging between 16°S and 40°N. About the temperature, *MRP2* and *MRP3* are present in a temperature range of about 0- 27°C; *MRM1* is also present in sea water with temperature of 30°C; while *MRP1* is not present in a temperature range of about 0-7°C and finally *MRM2* is not present in sea water with temperature greater than 20,8°C.

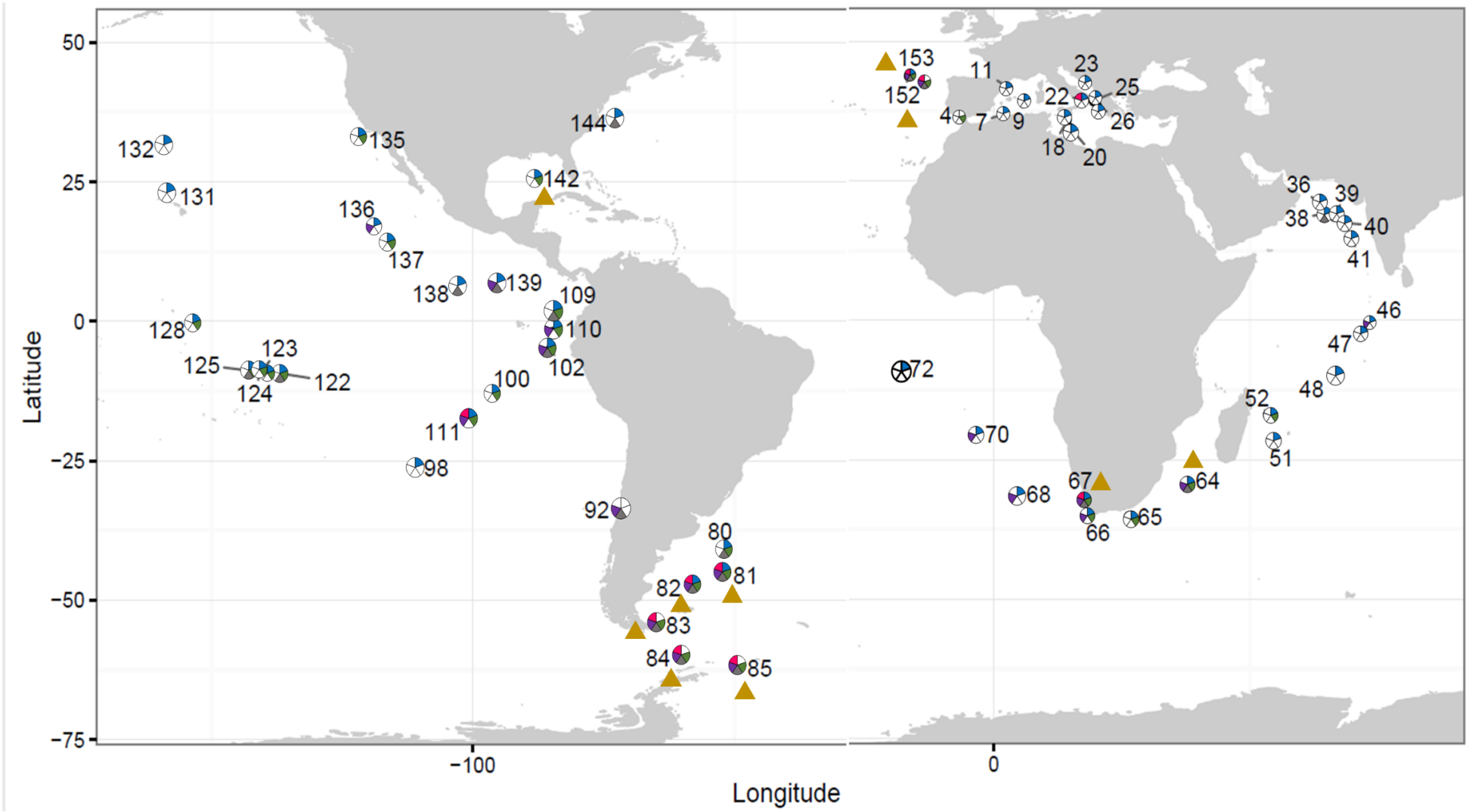


Fig. 3.11 Distribution pattern of the 5 MR genes and SPO11-2. For each station a pie chart shows presence (colour) or absence (white) of the MR genes, colours used for each gene are: blue for MRP1, green for MRP2, grey for MRP3, pink for MRM1 and purple for MRM2. Ochre triangle indicates the presence of SPO11-2

Tab.3.4 List of TARA stations for which information of species composition of the planktonic community is available (de Vargas et al. 2015; Malviya et al. 2016). The presence of *Fragilariopsis* and *Pseudo-nitzschia* genera and of MR genes are indicated. The circle represents presence of V9 metabarcoding sequences attributed to the genera *Pseudo-nitzschia* and *Fragilariopsis;* triangles represent observations in light microscopy. For stations 126 to 145 there were no public DNA metabarcoding data and the LM data were provided by E. Scalco.

		Genus		MR genes				
Station	T °C	*Fragilariopsis*	*Pseudo-nitzschia*	MRM1	MRM2	MRP1	MRP2	MRP3
4	18,3	O	O				x	
7	20,7	O	O			x		
9	20,2	O	O			x		
11	20,2	O	O			x		
16	15,6	O	O					
18	19,0	O	O			x		
20	21,5	O	O			x		
22	16,6	O	O		x	x		
23	15,9	O	O			x		
24	16,4	O	O					
25	16,8	O	O			x		
26	19,1	O	O			x		
30	20,2	O	O					
31	24,6	O	O					
32	26,2	O	O					
33	27,2	O	O					
34	27,6	O	O					
36	25,4	O	O			x		
38	26,0	O	O			x		x
39	26,9	Not found	O			x		
41	27,6	O	O					
42	28,3	O	O					
44	28,1	O	O					
45	24,0	O	Not found					
48	29,8	O	O			x		
52	26,4	O	O			x	x	
64	22,2	O	O	x		x	x	x
65	21,7	O	O			x	x	
66	15,0	O	O	x		x	x	
67	13,0	O	O	x	x	x	x	x
68	14,4	O	O	x		x		
70	19,8	O	O	x		x		
72	24,3	O	O			x		
76	22,6	Not found	O					
78	19,8	O	O					
79	17,6	Not found	O△					
80	18,9	Not found	O△			x	x	x
81	13,6	O△	O△	x	x	x	x	x
82	7,3	O	O	x	x	x	x	x
84	1,9	O	O	x	x		x	x
85	-0,3	O	O	x	x		x	x
93	16,8	Not found	O△					
98	18,4	O	O			x		
100	20,7	O	O			x	x	
102	22,4	O	O	x		x	x	x
109	22,5	O	O			x	x	x
111	20,8	O	O	x	x	x	x	
122	26,6	O	O			x	x	x
123	26,6	O	O			x	x	
124	26,5	O	O			x	x	
125	26,8	O	O			x	x	x
126	26,4	△	△					
127	27,0	△	△					
128	26,2	△	△			x	x	
135	14,7	Not found	△			x	x	
144	22,9	△	△			x		x
145	14,1	Not found	△					

	North Atlantic Ocean
	Mediterrenean Sea
	Red Sea
	Indian Ocean
	South Atlantic Ocean
	Southern Ocean
	South Pacific Ocean
	North Pacific Ocean
O	DNA metabarcoding data
△	Light microscopy data
x	Presence of gene
Not found	Genera not found
Empty cell:	Gene not found

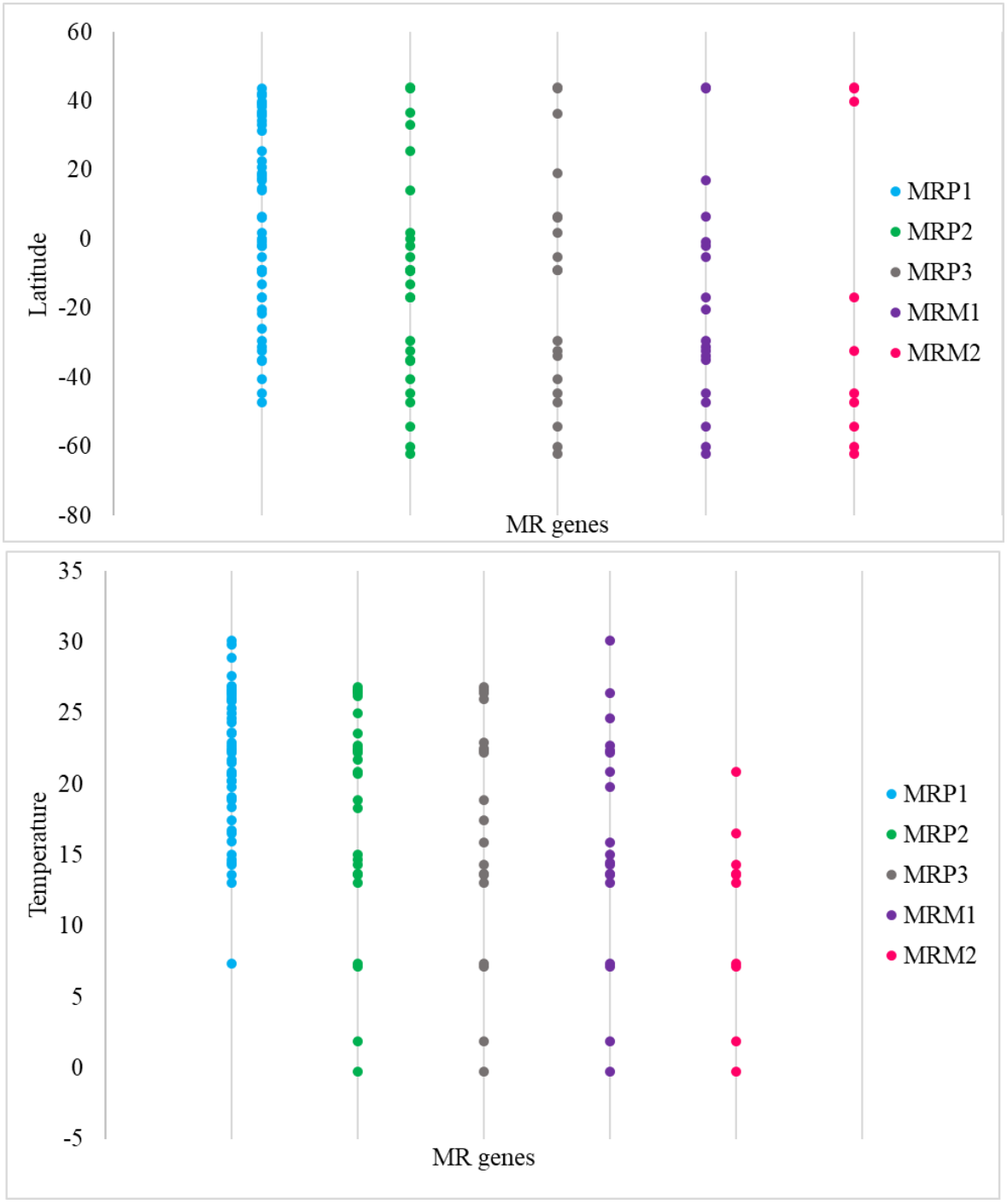


Fig.3.12 Distribution of MR genes in relation to latitude and temperature: top, distribution of homologues with respect to latitude; bottom, distribution of homologues with respect to temperature

3.3.4 Preliminary results of searches in the Tara metagenome

Only very recently, after I had performed the analyses presented in the previous paragraphs, the MATOU metagenome database (Tara Oceans Coordinators et al. 2018) became available (Villar et al. 2018). With little time available, I could only perform the first step of analysis (tblastn) to have a general overview in term of differences between the searches in metatranscriptome and metagenome (Table 3.5), even if this comparison has the limit of an untrimmed dataset. These rough comparisons highlight 3 different cases: i) genes for which the number of hits in both databases are more or less similar (*MRP1*, *MRP3*, *SPO11-2*), ii) genes for which the number of hits in the metatranscriptome is lower than that in the metagenome database (*RAD51-A*) and iii) genes for which the number of hits in the metatranscriptome is higher than that in the metagenome database (*MRP2*, *MRM2* and the striking case of *MRM1* for which no sequence could be retrieved in the metagenome).

Tab.3.5 Table with number of hits for searched genes in the metatranscriptome (MATOUv1+T) and in the metagenome (MATOUv1+G) databases.

Number of Blast analysis hits		
Gene ID	MATOUv1+T	MATOUv1+G
MRP1	54	56
MRP2	3196	1606
MRP3	7	5
MRM1	51	0
MRM2	7403	3455
SPO11-2	250	278
RAD51-A1	4554	7416

3.4 Discussion

The Tara metatranscriptome and metagenome databases represent an opportunity to explore over a wide spatial scale the pattern of distribution of genes involved in a specific biological process, in this case sexual reproduction.

The chosen genes were the MR genes, *SPO11-2* and *RAD51-A1*. The exploration of these genes in this data set was focused mainly on three points:

i. Defining the number of homologues of MR and meiotic genes
ii. Defining the spatial distribution of MR and meiotic genes
iii. Exploring the possibility to use these genes as molecular markers to detect sexual events during phytoplankton blooms.

In the following, I will discuss the results concerning these questions.

i) The phylogenetic analyses allowed to identify homologs for the MR genes; in particular, for *MRM1* there are 7 putative orthologues of which 3 assigned to *Fragilariopsis sp*, other that are likely of the same genus and one sequence that could belong to the *Pseudo-nitzschia* genus. Low conservation for this gene is in line with the data obtained searching in the sequenced diatom genomes and in the MMETSP database (Figure 3.2).

For *MRM2* and *MRP2*, despite the high number of hits due most likely to the presence of the common LRR domain, only 3 and 4 putative homologs, respectively, could be confirmed in the phylogenetic analysis. For *MRM2*, two could belong to the *Pseudo-nitzschia* genus and one identified as *Asterionellopsis glacialis*. Differently from the other MR genes, MRM2 appears to be present outside of the *Pseudo-nitzschia* and *Fragilariopsis* genera (one homolog detected in *Fistulifera solaris*, Figure 3.2), the finding of an *Asterionellopsis glacialis* homolog supports this notion.

For *MRP3,* 6 out of 7 TARA sequences displayed a good alignment, showing blocks of conserved amino acids especially towards the C-terminal part of the protein. *MRP3* was highly conserved in *Pseudo-nitzschia* and *Fragilariopsis* species (Figure 3.2) but not detected in other diatom genomes nor in any of the other MMETSP transcriptomes (Russo et al. 2018), comprising 178 assembled diatom transcriptomes from 92 diatom species cultivated under different conditions (Keeling et al. 2014). Despite only two of the six sequences had a clear assignation to either *Pseudo-nitzschia* or *Fragilariopsis*, TARA data support the fact that *MRP3* is not widely conserved in diatoms. MRP3 is involved in sex determination, a process known to be based on very different mechanisms also within the same groups (Bachtrog et al. 2014), therefore little conservation is not surprising.

For *MRP1* all 14 homologs present in the tree could be considered likely putative orthologues, but none of these can be traced back to known genera.

In Basu et al. (2017), given the evidence that proteins involved in sexual process are under positive selection (Clark et al. 2006), an analysis of the Ka/Ks ratio (number of non-synonymous mutations/number of synonymous mutations, indicative of the selection pressure on a gene sequence) was carried out. Among 434 genes, regulated during the early phase of sexual reproduction, 10 were under positive selection among which there was *MRP1* with Ka /Ks ratio equal to 2,13 (Table S15 in Basu et al. 2017), indicating a selective advantage to amino acid substitutions in the MRP1 protein, which is therefore quite variable in terms of primary sequence.

The search of MR genes has overall yielded few results in terms of reliable hits and in terms of assigned sequences. In terms of hits the results span from 7403 (*MRM2*) hits to 7 hits (*MRP3*). These broad differences could reflect the fact that, as already mentioned, some genes, like *MRM2* and *MRP2,* have common features such as a transmembrane region and a LRR domain. On the other hand, genes like *MRP3* and *MRM1* have low expression levels in the cell both during the vegetative and sexual phase (Russo et al.

2018). *MRP1* has no recognisable domains, except for the presence of a small putative signal peptide that targets proteins for translocation across the endoplasmic reticulum membrane (Russo et al. 2018), while *MRM1* has a Heat Shock Factor (HSF)-type DNA-binding domain. For both sequences about 50 hits were found, this number is very low considering the huge amount of data in OGA. Few results for *MRP1* and for *MRM1* could depend on their specificity for *Pseudo-nitzschia* and *Fragilariopsis spp)* and on abundance of these genera in the stations. The blast results of *MRM1* show alignments mainly for the region of the protein corresponding to HSF-DNA-binding, this domain should generate many more hits, also relative to other taxa, than those obtained. This result could be related to specific expression conditions of this genes.

In a recent study, it has been shown that in *Phaeodactylum tricornutum* iron starvation–induced protein 1 (*ISIP1*) plays an important role in the uptake of siderophores (high-affinity iron-chelating compounds). The authors proposed that it is part of an adaptive strategy that has allowed some diatom lineages to flourish in iron-limited environments (Kazamia et al. 2018). The *ISIP1* sequences were not found in prokaryotes, animals/fungi, viridiplantae, and eukaryotic algae, but were found only in diatoms; in particular, *ISIP1* was found in 28 of 99 diatom species (pennate and centric species, including the basally divergent genus *Corethron*). *ISIP1* was investigated also in the Tara Oceans metagenome and metatranscriptome where it was recovered like a diatom specific gene. The authors found a correlation between the gene expression abundance and distribution of *ISIP1* with diatoms in stations characterized by low iron concentration (Kazamia et al. 2018). The gene and transcript sequences of *ISIP1* were found in all 7 oceanic regions and the majority of the transcripts belonged to the orders of Thalassionemales (80%) and Chaetocerotales (14%). Also *Pseudo-nitzschia* sequences were found (Table 3.6).

Tab.3.6 *ISIP1* sequences for *Pseudo-nitzschia* from supplementary material (Table S1) of Kazamia et al. (2018).

Pseudo-nitzschia ISIP1 (Kazamia et al. 2018).		
Kind of meta data	Ocean province	Normalized Abundance
metagenome	Indian Ocean	0,00213423
metagenome	South Pacific Ocean	0,131713636
metagenome	South Atlantic Ocean	0,00542643
metatranscriptome	North Pacific Ocean	0,008238602
metatranscriptome	Southern Ocean	0,00242194
metatranscriptome	South Atlantic Ocean	0,008848349
metatranscriptome	South Pacific Ocean	0,235246571

In particular, 3 gene and 4 transcript sequences were found and in the oceanic region where these sequences were recovered *Pseudo-nitzschia* was present, but not always there was correspondence in the oceanic regions between recovered sequences. For example, 2 transcripts were found in North Pacific Ocean and in Southern Oceans, but gene sequences were not found in these oceanic regions. In view of this it is clear that there can be the case where the search of a gene in metagenome could be less efficient, even if it is a needed step to infer transcriptional differences among stations allowing to understand the regulation patterns of genes. Similarly, as for *ISIP1*, for *MRM1* in the very preliminary results of searches in the MATOUv1+G, genome sequences were not found, while transcripts were recovered, this case could be explained by the fact that genome content is lower than transcriptome content in a cell and by other technical problems like the nucleic acid extraction efficiency and the sequencing method.

ii) Looking at MR genes distribution in TARA (Figure 3.11) different patterns can be observed. The single maps (Appendix 2) show the distribution of homologs in TARA stations. At a glance, it is clear that some homologs are present in cold waters and at specific latitudes; so the stations, where the query genes were found, were plotted in graphs respect to their latitude and temperature (Figure 3.12). *MRP1* is widely distributed along latitude, but is never present below 47 ° S (TARA_083, TARA_084,

TARA_085), while the other MR genes are present until S 62° (TARA_85). The temperature in station TARA_82, where *MRP1* was detected, is equal to 7,3°C, while in the station TARA_85 the temperature is equal to -0,28 °C. The homologs of *MRM2* are not present at latitudes spanning from S° 39 to 16° N; this could mean that *MRM2* is mainly present in cold species, or, less likely, there could be a correlation between gene expression and temperature, and therefore a functional correlation with environmental conditions. Meta data for arctic stations are not yet available, these data could be useful to elucidate this hypothesis. In *Pseudo-nitzschia* sex occurs when opposite MTs meet and it is not triggered by environmental cues (Montresor et al. 2016), although environmental parameters might affect the efficiency of reproduction, or might favour a bloom where chances of encounters of a partner are higher. Further analyses have to be performed including the MR genes search in metagenome and exploiting the metabarcode data to investigate a possible correlation among species (metabarcode), genes (metagenome and metatranscriptome) and temperature/latitude. The metabarcode data, restricted to a set of stations covering seven oceanic regions, are available for all eukaryotic plankton (de Vargas et al. 2015). This data will be processed and trimmed to extract V9 metabarcode data of genera of interest and their relative abundance will be used for subsequent analysis aimed to understand a) if there is a correlation between abundance of V9 metabarcode of exanimated genera (*Pseudo-nitzschia* and *Fragilariopsis sp*) and abundance of recovered MR genes in MATOU, both meta-T and meta-G; b) what are the stations where, for the investigated genera, there are MR genes that were differentially expressed: for obtaining the differential expression data within a single station normalization steps like those performed for *ISIP*1 (Kazamia et al. 2018) will be required; c) finally a correlation between species/MR genes distribution with temperature and/or latitude could be attempted.

iii) *SPO11-2* and *RAD51-A1* were searched to understand if it is possible to consider a set of genes as a signature of sexual reproduction events at sea. Unfortunately, for

RAD51-A the phylogenetic resolution was not good enough to allow detection of the meiotic homolog, therefore this gene cannot be used to detect sexual events. *SPO11-2* is exclusively meiotic and highly induced in *P. multistriata* during gametogenesis (Patil et al. 2015). *SPO11-2* homologs were present in the stations 64, 67, 81-85, 142, 152 and 153 (Table 3.4), with higher abundance in stations 64 and 84. In 4 stations where *SPO11-2* was present also all MR genes were recorded, in 5 there were 4 MR genes and in the station 142 there were only *MRP1* and *MRP2*. Given that MR genes are not very widely distributed, the fact that *SPO11-2* transcripts often co-occur with MR transcripts is interesting. When sex is occurring, MR genes, except *MRP3*, are induced and it could be easier to see them all together. With the data available it is however currently not possible to assign genes to a given species and the presence of different MR genes could also be explained by the co-existence of different species each expressing one or a few MR genes. A taxonomic study of 46 TARA stations that cover seven oceanographic areas was conducted (Malviya et al. 2016) using V9. Metabarcodes revealed high diversity in diatoms groups and also in the genus *Pseudo−nitzschia* where more than 102 OTU were detected, mainly in the 5-20 µm size fractions. In this study the biological samples of 15 stations were examined in light microscopy to compare the morphological taxonomic identifications with the V9 identifications. In 14 stations the V9 and microscopy observations showed the presence of *Pseudo-nitzschia* and *Fragiliaropsis* genera. In these 14 stations I found at least 2 MR genes, and in particular in the stations 64, 82, 84, and 85 I found concomitant presence of MR genes and *SPO11-2*. The comparison between V9 and microscopy observations was not coherent for station 84: in this station the V9 metabarcode analysis predicted the presence of the centric diatoms *Chaetoceros* and *Proboscia* and the pennate *Pseudo-nitzschia*, while the morphological identification revealed the dominance of *Fragilariopsis sp*. In this station I found homologs of *MRM1* and *MRP3* both assigned to *Fragilariopsis sp*. and a homolog of *MRP3* assigned to *Pseudo-nitzschia sp*. This comparison among my data,

V9 and microscopy observations was not coherent also for station 82 (Malviya et al. 2016), in which *Fragilariopsis* was not detected by both V9 barcodes and LM, while I found *MRP3* and *MRM1* identified as *Fragilariopsis* homologs. In the metabarcode data of de Vargas et al. (2015) (see the krona of station 82, http://taraoceans.sb-roscoff.fr/EukDiv/) *Fragilariopsis* was recovered.

In vegetative cells MR genes have moderate basal expression levels, that increase, except for *MRP3*, during sexualisation, and among these *MRP1* is the gene most induced (Basu et al. 2017; Russo et al. 2018). So actually, although there could be a general upregulation of MR genes, the best gene for sex detection would be *MRP1* that was strongly induced during all phases of sexual reproduction (see results presented in the Chapter II). There are 19 stations in which only *MRP1* is found. In five out of ten stations in which *SPO11-2* was found, *MRP1* was also present (064, 067, 081, 082 and 153). Biological samples for these stations are available and should be examined carefully to search for the presence of sexual stages, which would support the use of these two genes in combination as markers. For station 102, in biological samples 4 auxospores of pennate diatoms were found (personal communication by E. Scalco), in this station 4 MR genes were found, but *SPO11-2* was not recovered. Auxospores appear a few days after the beginning of an event of sexual reproduction, and *SPO11-2* is specifically required during gametogenesis. A possible explanation could be that the sampling was performed too late in the process to capture *SPO11-2* upregulation. In RNAseq data relative to sexual reproduction up to 5 days (see Chapter II), this gene reached the peak at 24 hours (greater presence of gametes) and decreased at 5 days, being 3-fold less expressed than at 24 hours (data not shown).

To my knowledge this is the first attempt to determinate genes useful for sexual events in the marine environment. A similar attempt was done by Gong et al. (2017), in this work the authors try to select molecular indicators for bloom detection. The area of study was Neuse River Estuary (NRE) in North Carolina during the early fall of 2012.

The authors performed a comparative metatranscriptomic analysis of a dinoflagellate bloom area versus two stations with environmental characteristics that do not allow the bloom. From these comparisons they selected gene categories and proposed them like molecular indicators of bloom events. These were generic categories such as nucleotide sugar biosynthesis, other DNA/RNA processing, ATP synthesis, carbon fixation and photosynthesis. Furthermore, they proposed also genes involved in resisting cysts formation (mainly genes related to sterol synthesis) as indicators of bloom termination and so of the senescent phase (Gong et al. 2017). Differences between detection of sexual events and blooms in the sea exist due to the different magnitude of phenomena, sexual reproduction having a lower magnitude than a bloom. Furthermore, for detection of bloom events many gene categories can be used and these genes are common among species for their involvement in biochemical processes, while the genes relate to sexual reproduction, especially those involved in mate perception, are more diversified among species.

In this exploration, aimed also to propose a set of genes for detection of sexual events during diatom blooms, it emerged that *SPO11-2* could be used paired with MR genes. Suitable experiments possibly targeting specifically a *Pseudo-nitzschia* bloom should be designed to confirm the possibility to use the proposed genes as markers. Moreover, it will be necessary to expand the panel of genes considered, including other genes highly upregulated during the process, like for example the 7488 gene that I confirm as highly induced in Chapter II, and other genes covering all time windows of sexual reproduction. In the meantime, further studies on selected genes could be conducted exploiting TARA database and normalizing the data to obtain expression data of selected genes.

Chapter IV:

Metabolic differences between vegetative growth and sexual reproduction

4.1 Introduction

Marine organisms produce a huge variety of chemical compounds involved in several important processes like defence, intra- and inter- specific communication, sexual reproduction and feeding. These chemical compounds are called secondary metabolites, since they are not directly involved in the primary metabolism (Ianora et al. 2006).

Secondary metabolites include allelopathic compounds that inhibit the growth of other species. In diatoms, broken cells release polyunsaturated aldehydes and these compounds can be very toxic for the reproduction of their grazers, the copepods (Ianora et al. 2004), or can serve like signals for bloom termination through the induction of cell death (Vardi et al. 2006).

Metabolomics, the large-scale study of metabolites, has recently been applied to the investigation of the chemical signals employed by diatom cells in different processes. To explore the response of diatoms to the presence of grazers, RNAseq and metabolomics approaches were used, in particular comparing cultures of *Skeletonema marinoi* in presence of calanoid copepods with cultures without predators. A total of 242 regulated metabolites were identified and, among these, a set of downregulated compounds in the treatment with the grazers were annotated as oxylipins and, at the same time, RNAseq data showed downregulation of the pathway involved in oxylipins production (Amato et al. 2018). In another study 13 diatom cultures were grown in different nitrogen conditions and metabolomics analyses were performed. Different growth conditions affected metabolite accumulation but not the species-specific metabolomic profile, suggesting that the metabolic patterns is genetically determined and not simply modulated by nitrogen fluctuation. Even if was not consistent across all 13 diatom cultures, in low nitrogen condition an increase of triacylglycerol (TAG) was observed (Bromke et al. 2015).

Other important compounds for diatoms are sex pheromones, compounds that allow the species-specific recognition of opposite MTs and the synchronization for gametes production. Sex pheromones in diatoms are poorly known and they have been investigated only in one species.

As already mentioned in Chapter 1, the first sex pheromone characterized was the diproline, that is an attractive pheromone produced in *S. robusta* by MT- to attract MT+ (Gillard et al. 2013). Diproline elicited a significant attraction of MT+ cells at about 2 pmol mg^{-1} and a concentration-dependent attraction was observed with saturation at 200 nmol mg^{-1} (Gillard et al. 2013). D-diproline was active in a similar concentration of L-diproline. The authors hypothesized that there could be a "swipe card" recognition mechanism that is determined by electronic rather than geometric properties of a signal or in alternative it could be that racemization or other transformations generate an active product from D-diproline (Gillard et al. 2013).

Metabolomics has been applied to detect the chemical nature of *Seminavis* pheromones (Gillard et al. 2013). In general, it is conceivable that the major transcriptomic changes observed during sexual reproduction (Basu et al., 2017, Chapter II of this thesis) are accompanied by a parallel substantial rearrangement in the amount and nature of the metabolites produced by cells. As a complementary approach, I therefore explored the differences in metabolites profile between vegetative phase and sexual phase in the marine planktonic diatom *P. multistriata*. The metabolites produced or increased in the sexual phase could be putatively involved in pheromone production. The metabolite content of control MT+ and MT- was compared with the metabolite content of cultures in which sexual reproduction was occurring (cross). I sampled both the medium, for the exo-metabolome, and cell pellets for the endo-metabolome. Unfortunately, data relative to the exo-metabolome are missing due to technical problems during sample processing, and so only the endo-metabolome could be analysed.

For this analysis an ultra-high performance liquid chromatography (UHPLC) coupled with electrospray ionisation quadrupole time-of-flight (tandem) mass spectrometry (UHPLC-ESI-QTOF-MS(/MS)) was used to analyse samples. UHPLC is a method to separate individual constituents of complex natural product extracts, while with ESI-QTOF the sample is ionised by forcing a solution of the sample through a small heated capillary into an electric field to produce a very fine mist of ions, the time taken for each ion to reach a detector at a known distance is measured. The final result of extracted metabolites is a set of metabolomic 'features' (peaks) characterized by determined mass-to-charge ratio (m/z) and by a retention time, both unique for each compound, allowing the identification of compound thanks to comparison of these features against the in-house library (measured on the same instrument, with the same method) (Covington et al. 2017).

4.2. Materials and Methods

4.2.1 Experimental setup and sampling

For this experiment axenic cultures were obtained using a cocktail of antibiotics (final concentrations: 0.1 mg/ml streptomycin, 0.1 mg/ml penicillin and 0.5 mg/ml ampicillin). This experiment was aimed at characterizing metabolomics differences in two moments of the *P. multistriata* life cycle: the vegetative and the sexual phase. In particular, vegetative and cross samples were collected at four different time points to capture differences in subsequent phases of sexual reproduction. Specifically, samples were collected at 2, 6, 30 and 55 h (Figure 4.1), and for each time point three glass flasks were set up.

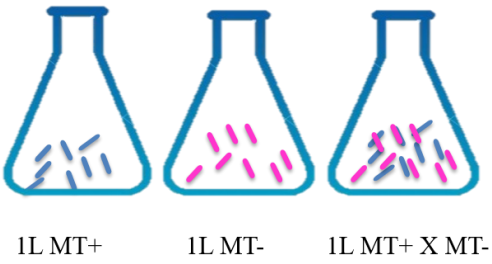


Figure 4.1. Scheme of experimental set up for the metabolomics characterization of the vegetative and sexual phases

Two strains were used: LV130 (MT+) and LV193 (MT-), both with an average cell length of about 30 µm. For each MT, cultures in exponential phase were synchronized in the G1 phase of the cell cycle with 36 hours of dark treatment; after synchronization the experiment was set up in a culture chamber at 18 ± 2 °C with a photoperiod of 12:12 hours Light: Dark and an irradiance of 100 µmol photon m^{-2} s^{-1}. A volume of 500 ml of each MT was put in co-culture in a glass flask (Cross), and 1 L of MT+ and 1 L of MT– were grown as monocultures in two separate glass flasks (Controls). When sexual stages were recorded in the cross, 800 ml of cell cultures were gently concentrated using Millipore® Stericup™ filters and then were centrifuged at 3220 g, 20 min, 18°C. The pellets were frozen with liquid nitrogen and stored at -80°C and lyophilized to obtain dry biomass. Dry pellets were weighed (Table 4.1). The samples were shipped to the laboratory of Mohamed Mehiri (University of Nice, France) for the production of extracts and then to the laboratory of Mark Broenstrup (Helmholtz-Zentrum für Infektionsforschung, Germany) for analysis.

Tab.4.1. Sample IDs and relative weight of dry pellets.

Sample IDs	Time points	Cultures	Dry weight
1X	2 h	Cross	10 mg
1+		MT+	10 mg
1-		MT-	10 mg
2X	6 h	Cross	10 mg
2+		MT+	10 mg
2-		MT-	3 mg
3x	30 h	Cross	10 mg
3+		MT+	10 mg
3-		MT-	30 mg
4x	55 h	Cross	16 mg
4+		MT+	16 mg
4-		MT-	12 mg

4.2.2. Biomass extraction protocol

3-50 mg of the dry pellets were processed to obtain two extracts by a mixture of MeOH/CH_2Cl_2 (50%:50%) to yield fraction F0 and by a mixture of H_2O/MeOH (80%:20%) to yield fraction Fw. The solvents used for extraction were removed from extracts by evaporation under vacuum (< 30 °C) and finally, the extracts were stored in the dark under nitrogen at -20 °C until the metabolomics analyses.

4.2.3. Metabolomic analyses

Chemical Fingerprinting of the extracts was carried out by Mohamed Mehiri'lab using high-performance liquid chromatography (HPLC) system with diode array detector (DAD), evaporative light scattering detector (ELSD) and Mass spectrometry (MS) (HPLC-DAD-ELSDMS) that allowed to determinate the number, the relative amounts, the chemical family and the molecular mass of the metabolites constitutive of the extracts, this part was. In Mark Broenstrup's lab the samples were processed in positive and negative ionization mode with an ultra-high performance liquid chromatography (UHPLC) coupled with electrospray ionisation quadrupole time-of-flight (tandem) mass

spectrometry (UHPLC-ESI-QTOF-MS(/MS)). UHPLC allowed physical separation of metabolites while the mass spectrometry allowed to determine the masses of molecules by ionizing chemical species and sorting the ions. Finally, data processing to obtain Based Peak Intensity (BPI) chromatograms, multivariate statistical analysis and annotation of compounds were performed.

4.2.4. Statistical analysis

Multivariate statistical analysis was based on data table with peaks and relative intensity in order to visualize the differences among the data sets. In this table were included only peaks detected with the centWave algorithm (Tautenhahn et al. 2008). Two most critical set parameters for centWave were: the peak width that is the expected range of chromatographic peak widths, given as a range in seconds (set to 10-30 seconds;) and ppm that is the maximum expected deviation of m/z values of centroids corresponding to one chromatographic peak (set to 15 ppm). The time at which compounds elute in the chromatography is different and it can vary in the different samples, so selected peaks (with centWave algorithm) were aligned: this step, called also retention time (RT) correction, aimed at adjusting retention time by shifting signals along the retention time axis to align the signals among different samples. Aligned peaks that matched among them were grouped in the features (Smith et al. 2006), but only chromatographic peaks present in at least 85% of the samples were used. Finally, these peaks were normalized using standards to remove technical biases and filtering to reduce the noise level in the data, so features with RT>=25 min and RT <= 0.8 min were removed. Statistical analyses performed were: principal component analysis (PCA) and orthogonal projection to latent structures-discriminant analysis (OPLS-DA). PCA is an orthogonal linear transformation of correlated variables into a smaller number of uncorrelated variables called principal components (PCs), where the greatest variance within the data is explained by any projection on the first coordinate (PC1) and the least variance is

projected by subsequent PCs (Jolliffe 2011). PCA displays as much variability (variance) as possible with as few PCs as possible.

OPLS-DA (supervised method), that respect to PCA gives a better class discrimination and more robust identification of important features (Trygg and Lundstedt 2007), was run to identify the features according to the order of contribution to the variation across the groups “vegetative samples” vs “Cross samples”. In the graph, T score axis represents the predictive variation among the groups and the Orthogonal T score axis represents the variation orthogonal to the group specific variation.

4.2.5. Annotation of features

Automatic Detection of Compounds was performed, and identification against the in-house library. Finally, metabolite research and identification of unknown features was performed using METLIN Metabolomics Database (Smith et al. 2005).

4.3. Results

The following analyses are data relative to the positive ionization mode acquisition (were the analyte is sprayed at low pH to have positive ion formation), data for the negative ionization mode acquisition were not available at the time of the study. The results are reported in two separate parts based on the solvents used for extraction: dichloromethane methanol (DCM) first and then water (H_2O).

4.3.1 Dichloromethane methanol (DCM)

To compare the metabolome of the samples (MT-, MT+ and Cross) at different time points, global profiling of the positive ion mode was analysed by UHPLC-MS. The Based Peak Intensity (BPI) chromatograms for each time point in positive ionization mode are reported (Figure 4.2).

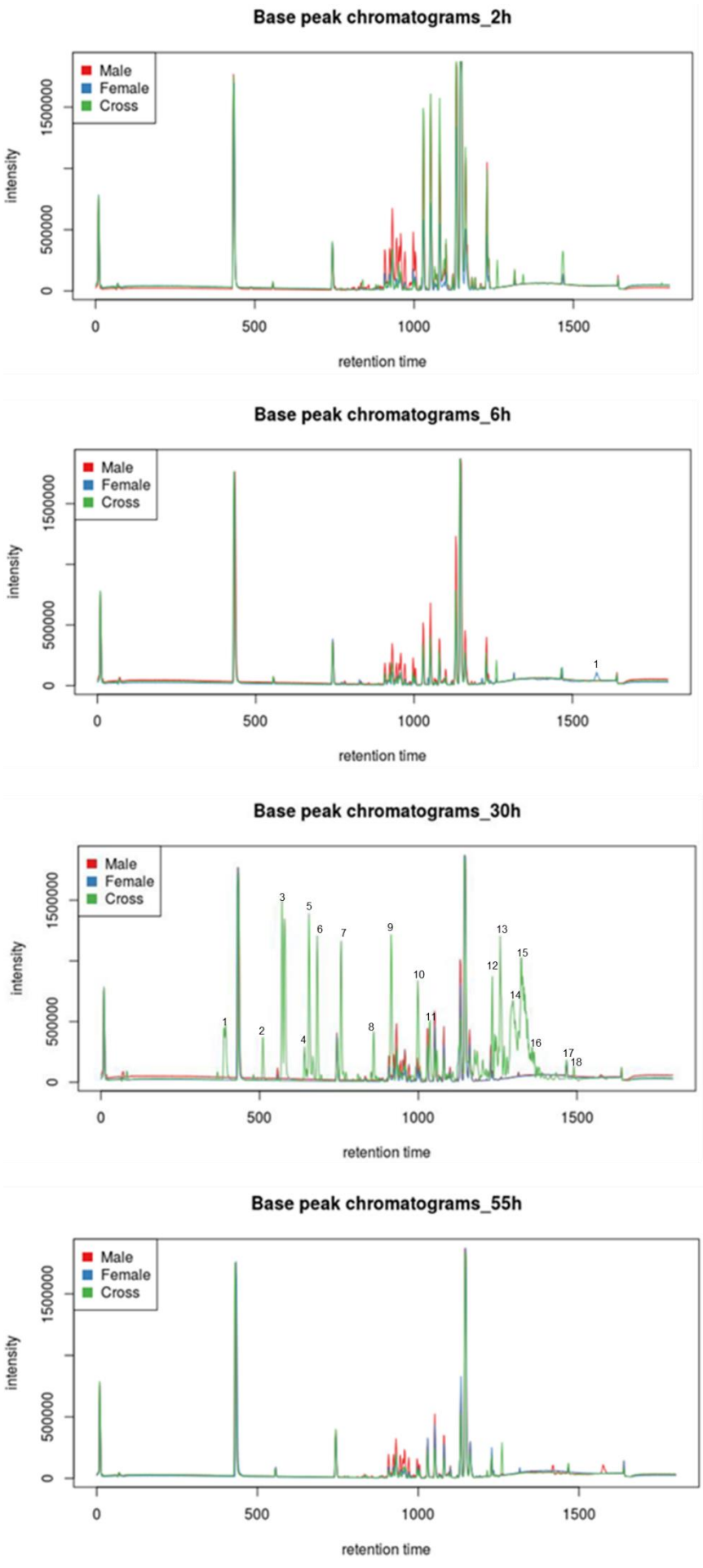


Fig.4.2. Based Peak Intensity (BPI) chromatograms for each time points comparing all samples (MT+ in red, MT- in blue and cross in green).

In general MT+ (male) showed different peaks and peaks of different intensity respect to the other two conditions (MT- and cross) at time points 2 h, 6 h and 55 h (Figure 4.2). At 6 hours a peak exclusive for MT- (female) was also observed (Figure 4.2, peak indicated with number 1). At time point 30 h the cross sample (green) showed additional peaks respect to MT+ and MT- (Figure 4.2, peaks indicated with numbers from 1 to 18). Normalized and filtered peaks were used to perform PCA (Figure 4.3) to have qualitative separation among samples. The samples scattering in the plots indicate similar metabolomic compositions when data are clustered together, while when metabolomic compositions are different samples are dispersed. The PCA scores plot divided the samples into different groups generally matching the two controls and the treatment, suggesting that the metabolic profiles have distinguished features in the MT-, MT+ and cross groups. However, the analysis also revealed a dispersion of some samples (MT+ 2h, MT- 2h and cross 2h).

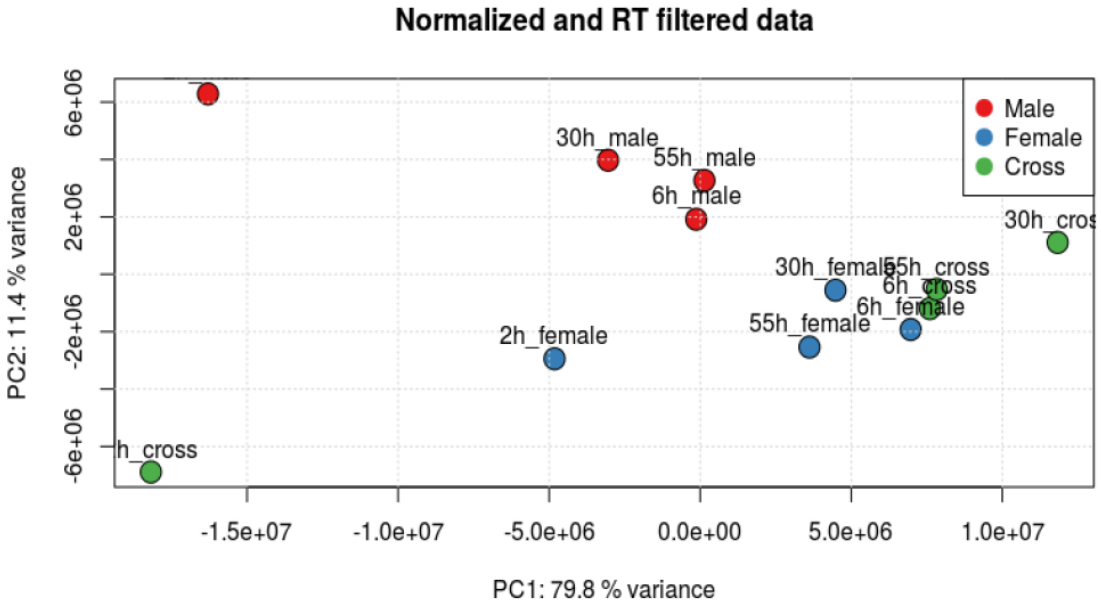


Fig.4.3. PCA using normalized and filtered peaks data. In the plot red dots indicate MT+ (male), blue dots indicate MT- (female) and green dots indicate Cross.

Since the aim was to identify differences between vegetative and sexual MTs (MT+ and MT- *vs* Cross) an additional analysis, that is OPLS-DA, was performed considering the two vegetative samples as one group and all the sexualized samples as another group (Figure 4.4). OPLS-DA allows a better class discrimination and more robust identification of important features. A statistical confidence region was added to the score plot as an aid in the detection of outliers.

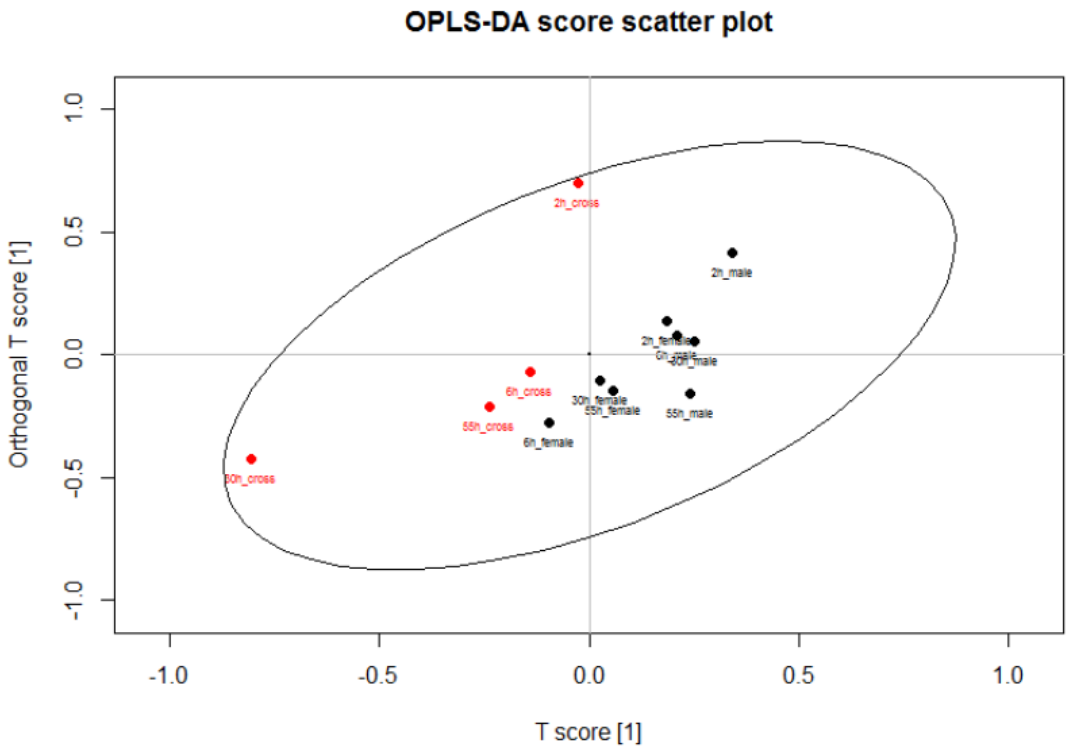


Fig.4.4. OPLS-DA using normalized and filtered peaks data separated in two main groups: black dots (MT+ and MT- samples) and red dots (cross samples).

Within the features that effected the separation presented in the OPLS-DA score plot, 15 were chosen (Table 4.2) considering their contribution to the variation and correlation within the data set.

Tab.4.2. The 15 most significant features that are responsible of separation of two sample groups in the OPLS-DA analysis. For each Feature it is shown the mass of the fragment (m/z), retention time in minutes (RT) and annotation (compound name).

Feature ID	m/z	RT (minutes)	Compound Name
M195T436	195,09	7,26	Phenazin-methosulfate
M228T1029	228,23	17,16	NA
M254T1054	254,25	17,56	**putative** palmitoleamide, C16H31NO
M256T1133	256,26	18,89	**putative** palmitic amide C16H33NO
M280T1084	280,26	18,07	Rac-glycerol 1-myristate
M280T929	280,26	15,48	NA
M282T1149	282,28	19,15	**putative** oleamide, C16H35NO
M283T1149	283,28	19,15	Petroselinic acid
M296T953	296,26	15,89	NA
M304T1149	304,26	19,15	NA
M310T1235	310,31	20,59	Erucic acid
M320T928	320,26	15,46	NA
M326T1037	326,38	17,29	NA
M332T1001	332,33	16,68	NA
M413T1263	413,27	21,05	Bis(2-ethylhexyl)Phthalate

Among these only 8 were annotated and for 3 of them the annotation was only putative. For each of these 15 features, the peak intensity is shown for the time points of controls and treatments (Figures 4.5-4.7).

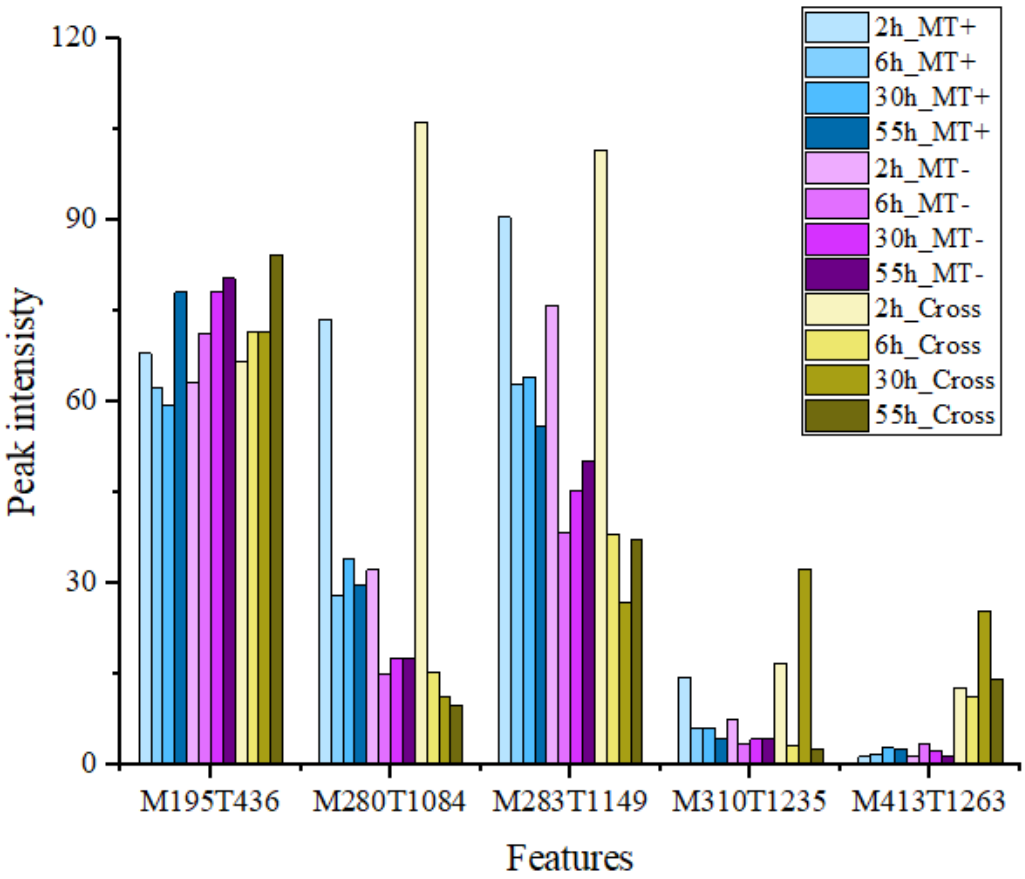


Fig.4.5. The five annotated features with their peak intensity in each sample for all time points. The peak intensity is divided for 100.000 for better viewing.

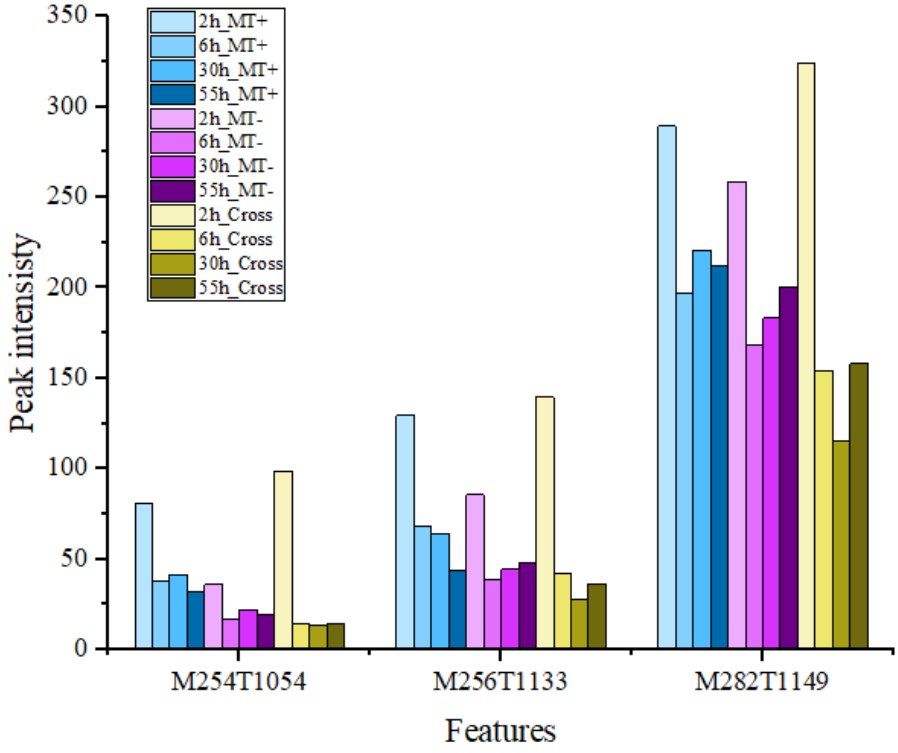


Fig.4.6. The three putatively annotated features with their peak intensity in each sample for all time points The peak intensity is divided for 100.000 for better viewing

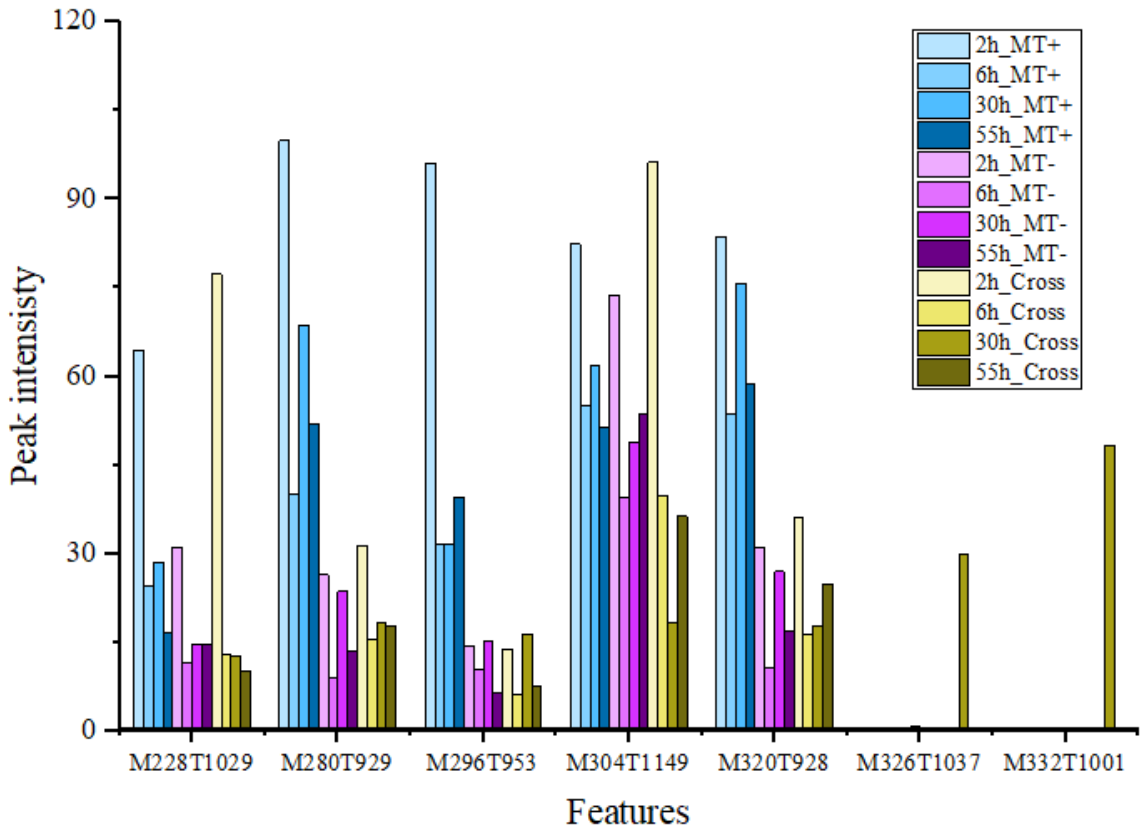


Fig.4.7. The seven unknown features and their peak intensity in each sample for all time points. The peak intensity is divided for 100.000 for better viewing.

4.3.2 Water extraction

The same analysis shown above for the DCM solvent is shown here for samples re-suspended in water. The Based Peak Intensity (BPI) chromatograms for each time point in positive ionization mode are reported (Figure 4.8).

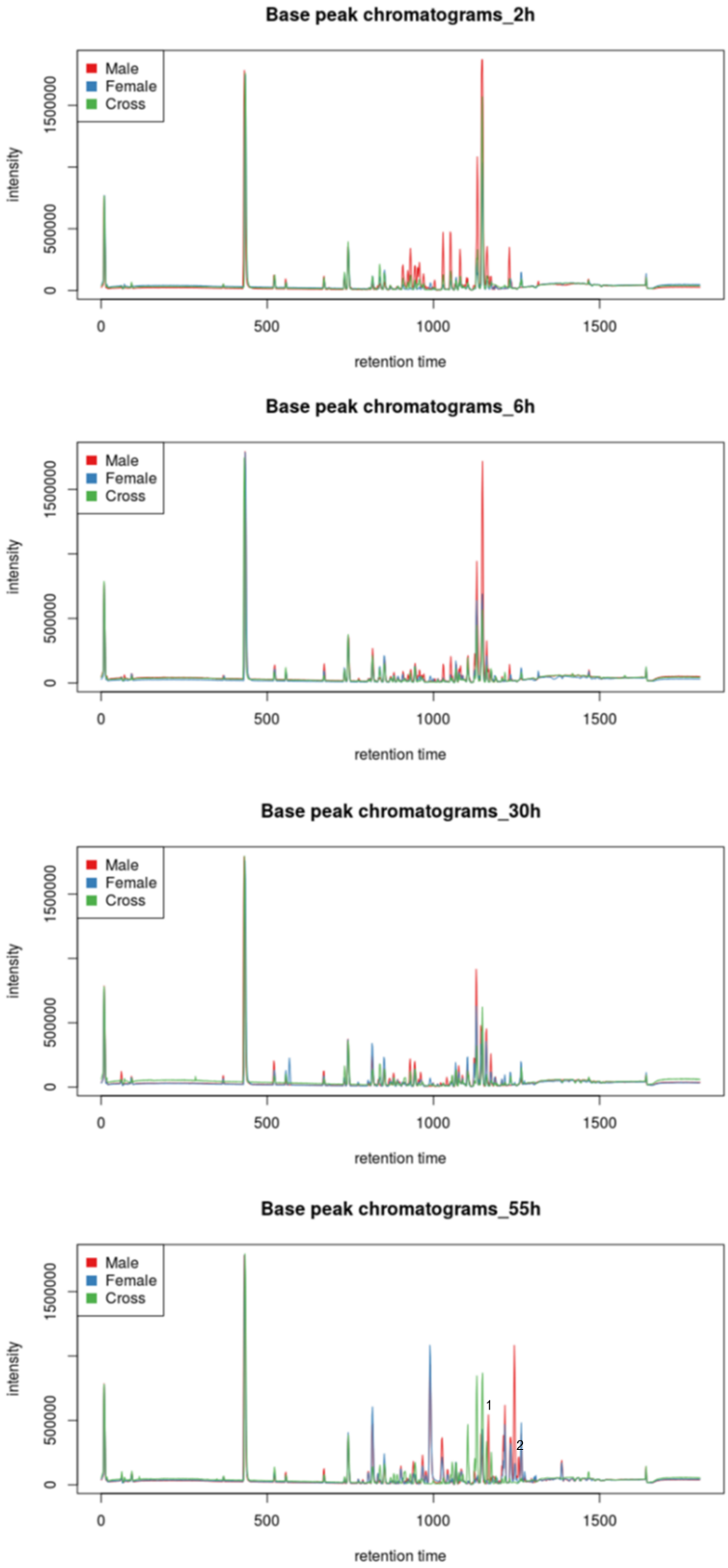


Fig.4.8. Based Peak Intensity (BPI) chromatograms for each time point comparing all samples (MT+ in red, MT- in blue and cross in green).

In general MT+ showed common peaks with different intensity respect to the other two condition (MT- and cross) at all four time points, while in the cross at 55 h there seemed to be two exclusive peaks with moderate intensity (Figure 4.8, peaks indicated with numbers 1 and 2). Normalized and filtered peaks were used to perform PCA (Figure 4.9) and OPLS-DA (Figure 4.10).

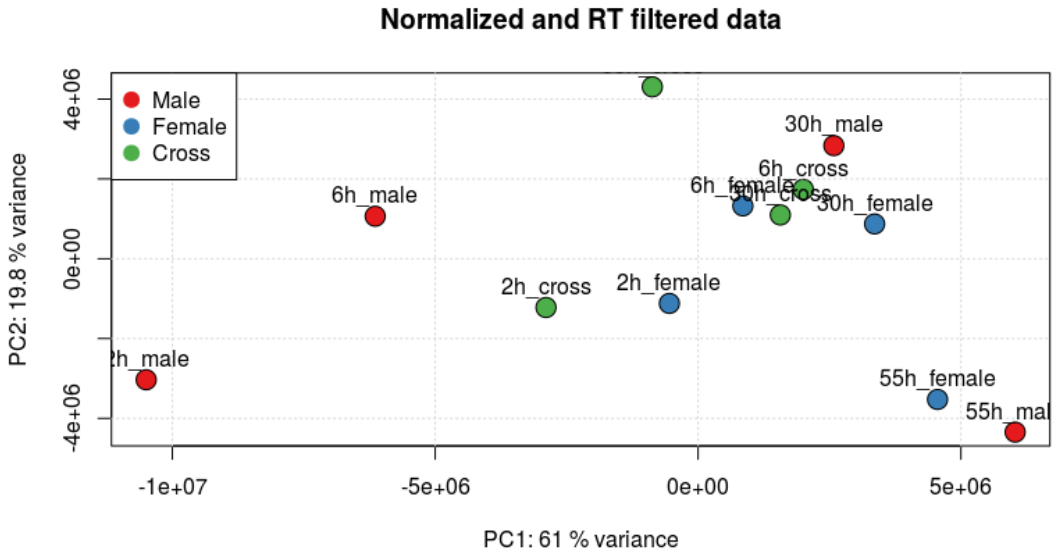


Fig.4.9. PCA using normalized and filtered peaks data. In the plot red dots indicate MT+ (male), blue dots indicate MT- (female) and green dots indicate Cross.

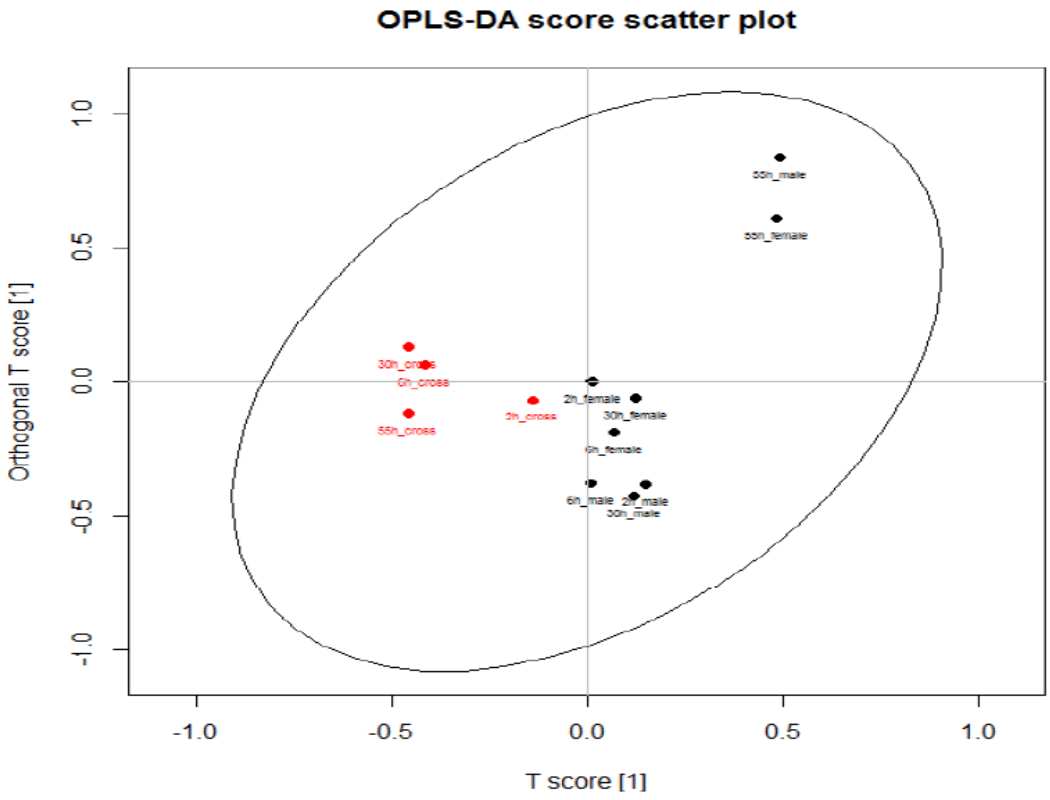


Fig.4.10. OPLS-DA using normalized and filtered peaks data separated in two main groups: black dots (MT+ and MT- samples) and red dots (cross samples).

Within the features that effected the separation presented in the OPLS-DA score plot, 7 were chosen (Table 4.3) considering their contribution to the variation and correlation within the data set.

Tab.4.3. The seven most significant features that are responsible of separation of two sample groups. For each Feature is shown mass of the fragment (m/z), retention time in minutes (RT) and annotation (compound name).

Feature ID	m/z	RT (minutes)	Compound Name
M195T435	195,09	7,26	Phenazin-methosulfate
M231T745	231,10	12,42	S-Naproxen
M494T1028	494,32	17,13	NA
M659T1146	659,43	19,10	NA
M724T1387	723,50	23,11	NA
M791T1234	791,47	20,56	NA
M792T1232	792,47	20,54	NA

Among these, only 2 were annotated. For each sample, a histogram of these features with relative peaks intensity over the time is shown (Figures 4.11-4.12).

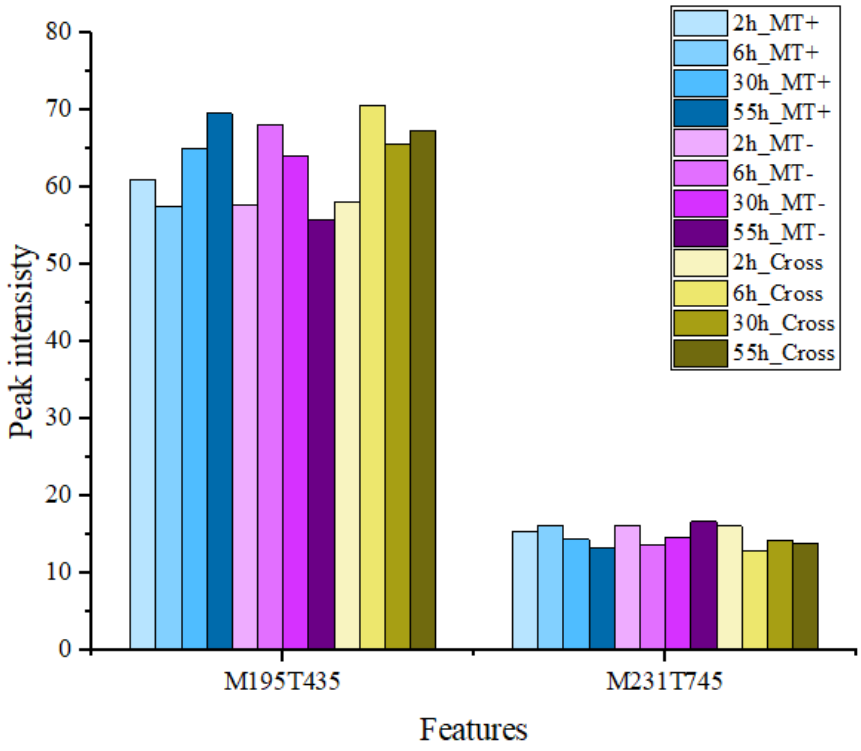


Fig.4.11. The histogram shows the 2 annotated features and relative peak intensity in each sample for all time points. The peak intensity is divided for 100.000 for better viewing.

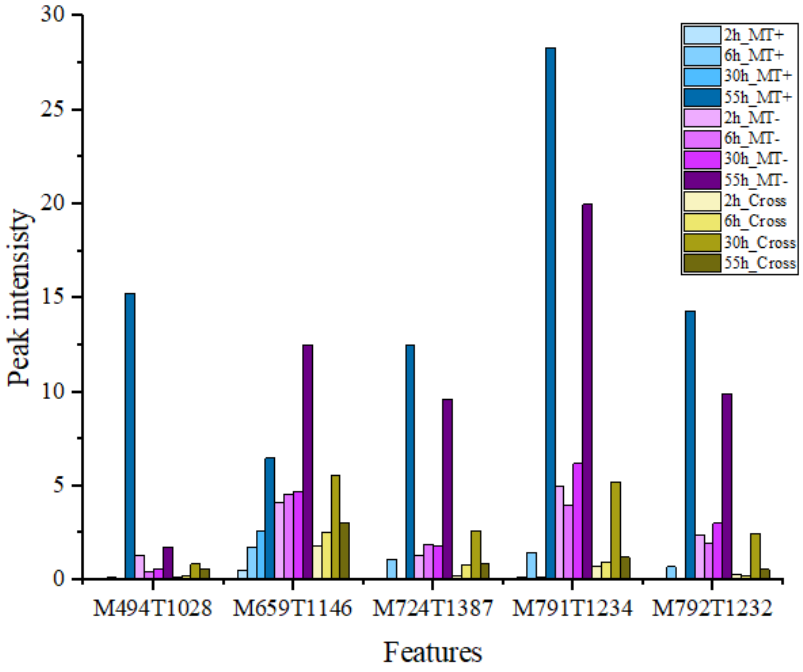


Fig.4.12. The histogram shows the 5 unknown features and relative peak intensity in each sample for all time points. The peak intensity is divided for 100.000 for better viewing.

4.4. Discussion

This first attempt to explore the metabolomics profiles of cells in the vegetative and in the sexual phase highlighted few differences in terms of metabolomics composition. Exclusive compounds discriminating MTs were not detected, while two unknown compounds exclusive of cross at 55 hours were identity (Figure 4.7). In total 22 compounds were identified differing among the two groups of samples for their peaks intensity. Nine out the 22 present annotation, they are: phenazin-methosulfate, palmitoleamide, palmitic amide, Rac-glycerol 1-myristate, oleamide, petroselinic acid, erucic acid, Bis (2-ethylhexyl) Phthalate and S-Naproxen.

Among these annotated compounds, those that presented high peaks intensity in the cross samples were: glycerol 1-myristate, oleamide, petroselinic acid, erucic acid and Bis (2-ethylhexyl) Phthalate.

Glycerol 1-myristate (PubChem CID: 79050).is a monoglyceride known as a monoacylglycerol, it is a glyceride constituted of one fatty acid chain covalently bonded to a glycerol molecule through an ester linkage. Monoacylglycerols are present in plant and animal tissues at very low quantities, and indeed their chemical nature would have a disruptive effect on membranes, having these molecules strong detergent properties. For their chemical features, these compounds are commonly added to commercial food products acting as emulsifiers, making ingredients, such as oil and water, blended (PubChem CID: 79050). This compound peaked at 2 hours in the cross, even if the release of gametes does not occur at 2 hours this compound could be accumulated for membrane disgregation and consequent release of gametes.

Oleamide is an amide of the fatty acid oleic acid that was found in all growth phases of *Chaetoceros gracilis* (Pratiwi et al. 2009). Oleic acid is a precursor of long chain polyunsaturated fatty acid (PUFA) (Pratiwi et al. 2009). Oleamide is an endogenous substance that accumulates in the cerebrospinal fluid during sleep deprivation and

induces sleep in animals. There is clinical interest for this compound as a potential medical treatment for mood and sleep disorders but also for depression caused by cannabinoid (PubChem CID: 5283387). Oleamide was found in the filamentous bacterium *Leucothrix mucor*, this substance reduces spore settlement and germling development in the fouling green seaweed *Enteromorpha pertusa* (Cho 2012). The oleamide, the major component of the *Mytilus edulis* periostracum, reduces the settlement and germination of the spores of the red seaweed *Pyropia suborbiculata* (Kang et al. 2016). This compound was tested also for suppression of *Pseudo-nitzschia pungens* that is capable of forming films when grown in plates. High concentrations of oleamide in the wax plates progressively reduced the attachment of this species (Getachew et al. 2016).

Petroselinic acid is the cis-isomer of octadec-6-enoic acid, it has a role as a plant metabolite (PubChem CID: 5281125). It is an important oleochemical compound for the food, cosmetics, chemistry and pharmaceutical industry.

Erucic acid is a monounsaturated omega-9 fatty acid found mainly in the Brassica family (PubChem CID: 5281116). Both compounds were found in the diatom *Chaetoceros gracilis* during growth, in very small amounts and only during the stationary phase. Palmitic amide is a fatty acid amide that derives from palmitic acid, a saturated fatty acid (SAFA), that was found also in all growth phases of *C. gracilis* (Pratiwi et al. 2009). This SAFA is always present in high quantity in Bacillariophyceae, and it is associated to the energy storage requirement, mainly during the later growth stages for cells division (Tonon et al. 2002).

Palmitoleamide is related to palmitic amide, but in its chemical structure it has two missing atoms of H.

Bis(2-ethylhexyl)phthalate (DEHP) is an ester of phthalic acid, it is used for the production of polyvinyl chloride (PVC) (PubChem CID: 8343). Curiously, in literature is reported that in the red alga, *Bangia atropurpurea*, cultured in artificial sea water

medium had similar amounts of DEHP to those cultured in natural sea water medium. The authors hypothesized that this alga synthesizes *de novo* phthalate esters (Chen 2004). Furthermore, since they found that DEHP contents in different species of algae grown in the same environment were different they asserted that this difference was due to the intrinsic nature of algae (Chen 2004).

Phenazines are heterocyclic compounds containing nitrogen produced exclusively by bacteria. Phenazines derived from the shikimic acid pathway, a highly conserved pathway involved in the production of numerous metabolites necessary for primary growth of bacteria. These compounds are used as electron shuttles, and are involved in cellular redox states, in gene expression regulation, biofilm formation and help bacterial survival (Pierson and Pierson 2010).

Naproxen is a propionic acid (PA) derivative with anti-inflammatory activity. PA is present naturally at low levels in dairy products and occurs ubiquitously, together with other short-chain fatty acids (PubChem CID: 1032). PA is reported as sex pheromone in the pherobase (http://www.pherobase.com/database/compound/compounds-detail-3Acid.php?isvalid=yes).

PA was found in oogoniol (pheromonal mixture) of species of *Achlya*, a genus of Oomycete (see Chapter two in Agosta et al. 1992).

Even if from statistical analysis in total 22 features resulted to differ in peak intensity among samples, only 8 showed marked difference in terms of quantity, these were compounds extracted with the DCM solvent. Among these five were unknown compounds, three of which with the highest peak intensity in MT+ (M228T1029, M280T929, M320T928), while the other two were exclusive of cross at 30 hours (M326T1037 and M332T1001). The other three features were annotated, they are: rac-glycerol 1 –myristate, erucic acid and DEHP. Rac- glycerol 1-myristate had the highest peak intensity at 2 hours in the cross, but this compound was in any case high also in the other samples. Erucic acid was present in major quantities in crosses at 2 and 30 hours,

while DEHP had the highest peaks intensity in all crosses, while in all vegetative samples it was very low.

These results provide novel data on compounds produced by *Pseudo-nitzschia*. Data will become more robust by using biological replicates for each of the specific conditions: in the analysis presented here we compared two biological groups (vegetative and sexualized MTs) to look for clear signals distinguishing the sexual phase, while triplicate will allow to detect/confirm compounds specifically produced or upregulated very early (putative pheromones), in intermediate phases (possibly compounds involved in cell adhesion, gamete release, and so on) and late phases (compounds produced by the large new generation cells). Moreover, in the same experiment illustrated here pellets for RNA were also collected and are available to explore the expression of genes putatively involved in the biosynthetic pathway of the compounds identified.

Any of the known compounds, if commercially available, could be obtained and administered to the cultures to verify possible effects on the efficiency of the sexual phase.

These data have been produced within a European project, the European Marine Biological Research Infrastructure Cluster (EMBRIC). In EMBRIC, one of the goals is to identify novel compounds from marine microalgae, and for this reason several diatom cultures have been processed with the same method used for my experiments. I participated also providing biomass for two more *Pseudo-nitzschia* cultures, while other teams provided cultures of other pennate diatoms with a set up similar to the one I used. Data from the project have not been fully processed but once they become available it will be possible for me to compare the metabolic changes occurring in *Pseudo-nitzschia* during sexual reproduction with the metabolic features of other diatoms, to look for common and unique metabolites.

Chapter V:

General conclusion and future perspectives

Phytoplankton is a key component in aquatic ecosystems. Traditionally, the dynamics of phytoplanktonic communities have been considered mostly dependent on environmental conditions, especially light and nutrient availability (Falkowski 1998; Falkowski and Oliver 2007). Indeed, biotic and endogenous factors are similarly important in the regulation of species successions and dominance. Phytoplankton life cycles are inevitably linked to population dynamics, still for many of the major groups little information is available, with limited understanding of the mechanisms regulating them and scarce knowledge of the molecules involved.

My PhD studies were focused on a planktonic diatom, *Pseudo-nitzschia multistriata*, and were aimed at elucidating molecular aspects related to a key phase of sexual reproduction: the mate perception between opposite MTs. This is interesting because it is still completely unknown how *P. multistriata* cells, which drift passively and are incapable of swimming, find and pair with their mating partners, and also because in general the mechanisms underlying the perception of external signals in diatoms are still poorly defined.

For *Pseudo-nitzschia multistriata*, genes involved in the early phase of sexual reproduction and in the mate determination system were identified in previous studies (Basu et al. 2017; Russo et al. 2018). During my work I focused on a set of genes selected for their putative involvement in the molecular cross talk that allows opposite MTs to recognize each other and to trigger meiosis. The selected genes were investigated in several experimental conditions to reconstruct the dynamics and hierarchical interaction between MTs.

The mate perception dynamics and the nature of the pheromones involved were elucidated for pennate benthic species *S. robusta* (Gillard et al. 2013; Moeys et al. 2016), but no pheromone is known for planktonic species, and one important goal of

my thesis was to find a method to allow the identification and isolation of *P. multistriata* pheromones.

Despite a high level of noise in my gene expression studies, I have identified a gene to use as molecular marker for testing the medium of MT+ in a bio-assay guided fractionation. The identified molecular marker (ID 7488) is upregulated when MT- perceives the medium of MT+, indicating that MT- perceives the MT + pheromone and this has an immediate effect on cell cycle regulation, as evident from the downregulation of the cyclin *CYC1B*. This is similar to what happens in *S. robusta*, in which the perception of SIP+, the constitutive pheromone of MT+, determines the switch from mitosis to meiosis of MT- cells (Moeys et al. 2016).

One of the hypotheses that can be made based on past and current data is that the constitutive putative pheromone of MT+ perceived by MT- could be *MRP1*, a gene, encoding for a small secreted protein, that is differentially expressed between opposite MTs and then highly induced during sex in the MT+. *MRP1* was not systematically induced in MT+ treated with the MT- medium, possibly because the *MRP1* expression could require multiple molecular signals from MT- or growing levels of the first MT- cue. To distinguish between these hypotheses, vegetative MT+ cells could be treated with the medium of sexualized MT-.

Following gene expression over time in the very early phase of sexual reproduction, an interesting finding was that the general upregulation of the MR genes in sexualized cells was higher in the experiment where *MRP1* had the highest expression level in the vegetative MT+ culture. *MRM2* during time courses in one experiment was induced late, the delayed induction in sexualized cells could be related to higher expression levels of *MRP1* in MT+ cells (see exp013 in Chapter II): *MRM2* expression levels could be therefore linked to MRP1 levels.

In the experiments where crosses were set up using different sex ratios a low number of mating events (pairs and clusters) was observed in 10:90 MT+/MT-, this is also an indication that the MT+ strain (perhaps through MRP1 levels) influences the reproductive efficiency. On the other hand, a high number of sexual events observed in the 10:90 MT-/MT+ ratio might suggest that there must be an MT- primary cue with powerful signalling capabilities that induces sexualisation in MT+.

MRP3 overexpression in an MT− strain induced sex reversal, the transgenic MT- in crosses with wild type MT− strain generated sexual stages. In the transgenic MT- also the expression of the MR genes was inverted, indicating that the regulation of these genes is *MRP3* dependent (Russo et al. 2018). The downregulation of *MRM2* in MT+ cells could be directly due to *MRP3* or could depend on *MRP1* which is turned on by *MRP3*.

These results indicate that, most likely, both MTs have a constitutive pheromone. *MRP1* is the candidate for MT+, and the gene with ID 7488 could be the counterpart in the MT- cells, since it displays the same behaviour: it is already differentially expressed between the two MTs in co-culture, and it is highly induced in the MT- upon sexualisation. Based on my results, 7488 should be considered the sixth MR gene and should be called *MRM3*. In the model of interaction between MTs, the MT+ is sexualized after sexualisation of MT- by a putative pheromone that could be 7488, alias *MRM3*, and future work will be aimed at testing this possibility.

To provide support to the idea that 7488 and *MRP1* genes could encode signalling molecules, *MRP1* could be produced in a heterologous system, it could be used to treat MT- cells, and to verify if it is effective I could follow the expression of 7488. Furthermore, if one hypothesizes that the pheromones are of proteic nature, the protease trypsin could be used to cleave putative cues contained in the medium, the

MT+ medium treated with trypsin could be used to treat MT- cells and to assay the expression of 7488.

In addition, to elucidate the role of 7488, knockdown transgenic lines could be produced for functional studies aimed to verify the phenotype during sexual reproduction in crosses.

As far as the novel information gained about signal transduction mechanisms, the upregulation of soluble guanylate cyclase (sGC) gene during sexual reproduction (Basu et al. 2017) indicates that cyclic guanosine monophosphate (cGMP) is involved in the process. From preliminary results (Chapter 2) *MRP1* induction should be independent of a putative signalling pathway mediated by *sGC* that is activated by nitric oxide produced by the nitric oxide synthases (NOSs). To confirm this hypothesis MT+ cells could be treated with 8-Br-cGMP, this compound is a cGMP analogue and it is cell permeable with higher resistance to hydrolysis by phosphodiesterases than cGMP. If the induction of *MRP1* is independent from activation of *sGC* its expression should not increase in the vegetative cells treated with 8-Br-cGMP. In parallel the same treatment could be done on vegetative MT- cell to check the expression of 7488.

Given that the work on diatom sexual reproduction has now provided a good deal of information about the genes involved, I ventured into the exploration of metagenomic datasets to verify whether it is possible to demonstrate that sexual events at sea are linked to the presence of specific transcriptional signatures. The opportunity to collaborate with SZN researchers who have been heavily involved in the Tara Oceans exploration allowed me to work with novel datasets that contain precious information. The MR genes, being the genes induced in sexual reproduction in *P. multistriata* and being present in the *Pseudo-nitzschia* and in the *Fragilariopsis* genera, were selected as

potential molecular markers to use for detection of sexual reproduction at sea, in combination with two meiosis related genes, *SPO11-2* and *RAD51-A1*. While more work will be needed for the RAD51 family, the results obtained with the other genes are encouraging: in few stations there was co-occurrence of all or 4 MR genes with *SPO11-2* indicating that perhaps sexual reproduction was going on in those stations. To refine the data, it will be necessary apply a normalization step using: 1) the abundance of OTUs related to species of interest; 2) the presence of MR genes and SPO11-2 in the metagenome; 3) the presence of MR genes and SPO11-2 in the metatranscriptome and 4) the total amount of transcripts assigned to species of interest.

This effort is useful to propose a set of genes to test for detection of sexual reproduction at sea in a suitable experiment during a *Pseudo-nitzchia* bloom. Sexual reproduction of *Pseudo-nitzschia* was rarely recorded in the natural environment and only two publications report these events: sexual stages were observed during a bloom with cell density ranging from 187,000 to 91,000,000 cell/L of *P. australis and P. pungens* in the NW Pacific coast (Holtermann et al. 2010), and *P. cf. delicatissima* and *P. cf. calliantha*, in the Gulf of Naples, Mediterranean Sea (Sarno, Zingone, and Montresor 2010). Actually, *Pseudo-nitzschia* species were often recorded in thin layers with cells density ranging between 50,000 and 30,000,000 cell/L, in particular colonies of *P. cf. pseudodelicatissima* were associated for 3 weeks with colonies of *Chaetoceros socialis* (Rines et al. 2002). Thin layers can be thick up to about 5 m and can spread horizontally for kilometres, this structure of colonies can dissolve by physical factors like wind or water density, but a permanent thin layer represents a good opportunity for mating, facilitating chemical communication or increasing the likelihood of cell-to-cell encounter, leading to sexual reproduction (Lelong et al. 2012). Sometimes events of sexual reproduction could occur in isolate spots in the thin layer, so the use of these genes for sexual detection could fail, while the approach might be more successful with higher cell densities as in the case of blooms.

The consultation of TARA metatranscriptome permitted also to study the global distribution of MR genes; as expected, MR homologs were found in *Pseudo-nitzschia* and *Fragilariopsis sp*, (Russo et al. 2018) but the spatial distribution of these homologs was not uniform. For example, MRP1 homologs were detected in all oceanic regions while MRM2 genes were mainly present in cold stations. It is likely that the detection on a global scale of *MRP1* transcripts could depend on its high constitutive expression, while the localized presence of *MRM2* in cold stations could mean that this gene is mainly present in cold species, or less likely that it is linked to temperature. The distribution of these genes on a global scale has to be studied browsing the metagenome, and possible inferences about expression of these genes require steps of normalization, considering also the abundance data of OTUs. The metagenome could also reveal new homologs not present in the metatranscriptome.

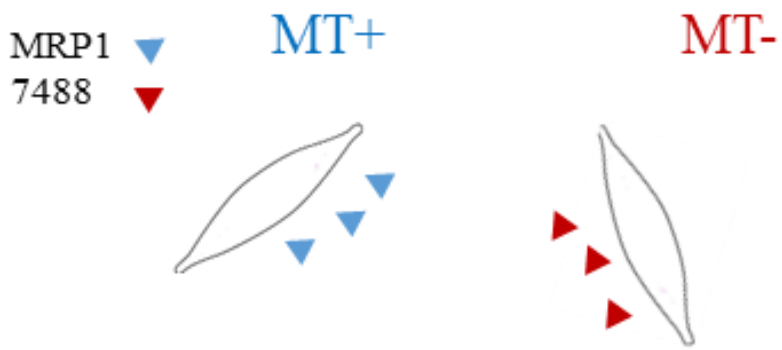
This exploration in TARA, aimed to respond to several questions, highlighted the potential of this tool, but also the limitations and problems. Sometimes there were discrepancies between my results and information relative to presence of genera at specific stations, this limit could be overcome by using additional metabarcoding, since sometimes the resolution at genus levels of V9 was low (Malviya et al. 2016).

In the context of this thesis I was interested also into discovering metabolomics differences that characterize the vegetative and sexual phases of the *P. multistriata* life cycle, since the metabolic profile of these two phases could lead to the discovery of signalling compounds in diatoms and could also revel new interesting compounds for human applications. The metabolomics analysis revealed that there were not exclusive metabolites distinguishing vegetative cells and sexualized cells, except at 30 hours in the cross (when parental cells, gametes and zygotes were mainly present) in which two unknown metabolites seemed to be exclusive. The validation of exclusive compounds of one specific time point will need replicates. Curiously, in the cross in each time

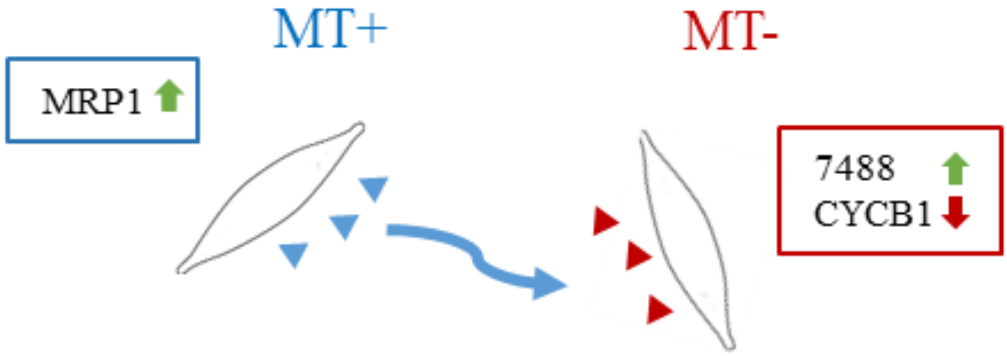
point, the peak intensity of the “feature” annotated as DEHP was the highest respect to vegetative cells. Other metabolomics experiments have to be performed using biological replicates and furthermore it would be interesting to analyse the metabolites released in the medium during the early phase of sexual reproduction to have a picture of the pheromones released in the medium after mate perception.

The results reported in this thesis provide novel information on several aspects concerning the dynamics of interaction of the two sexes of a planktonic diatom, elucidating part of the intracellular signals and hierarchical activation of genes. The supposed interaction between opposite MTs admits the existence of two constitutive pheromones, perhaps encoded by *MRP1* for MT+ and 7488 for MT-. These two genes are strongly induced during sexualisation respectively in MT+ and MT-, and in particular the first MT to sexualize is MT- that perceiving MT+ induces the transcription of 7488. A possible model could be that 7488 is the signal perceived by MT+ and that this signal is needed for induction of *MRP1* (Figure 5.1).

1) Both MTs have constitutive pheromones: MRP1 and 7488



2) The perception of MT+ triggers the induction of 7488 and the downregulation of CYCB1. The expression of MRP1 is induced



3) The perception of sexualized MT- strengthens the induction of MRP1, already induced perceiving MT-

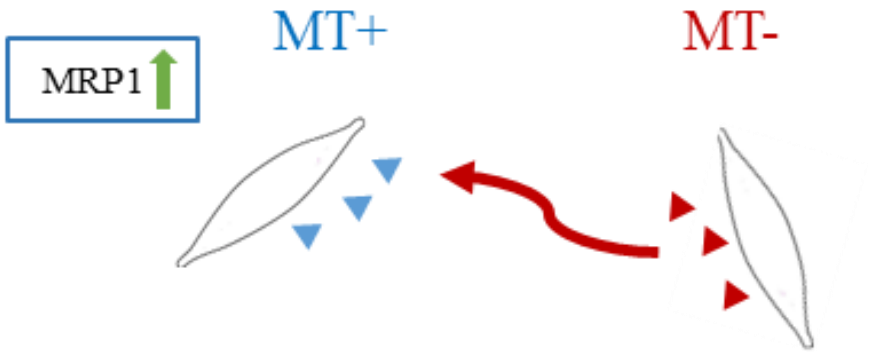


Fig.5.1. One of the possible models of interaction between the opposite MTs during sexual reproduction.

Since the gene 7488 was strongly induced when MT- perceived cues present in the MT+ free medium, it is now possible to set up a molecular bioassay for testing medium fractions with aim to isolate and characterize the putative constitutive pheromone of MT+; furthermore, the identification of the MT+ cue will allow to explore more clearly MT- gene expression changes in response to the first MT+ signal using RNA-seq, as already done in *S. robuta* (Moeys et al. 2016). Metabolomics profiles of the vegetative cells and cells in sexual reproduction identified a small number of compounds that appear to be enriched or exclusive for the sexual reproduction phase, however due to the lack of annotation for many of the genes as well as for features in the metabolomics data it is currently not possible to link the gene expression data with the secondary metabolites content. As reference databases constantly improve, it will be possible in the future to …

From an ecological point of view, this work indicates that genes related to sex could be makers for sexual detection at sea, and the future prospective is to extend this analysis to other genes belong to centric species and validate their use in the field monitoring diatom blooms in the LTER-MC station in the Gulf of Naples. Finally, the knowledge about sexual dynamics of planktonic diatom species could be useful to infer part of the unknown processes governing the dynamics of planktonic community.

Appendix 1

Abundance of homologs

The homolog abundances for each MR gene in all stations where the homologs were found.

			MRM1						
			Bacillariales	Fragilariopsis	Bacillariophyceae	Bacillariophyceae	Fragilariopsis	Fragilariopsis	Bacillariaceae
Stations	Samples	Fractions	81048584	113884330	111952663	112417579	96695084	113658669	116156696
46	N000000267	0.8-5µm	0,000002718447%						
64	N000000531	5-20µm					0,000006546605%		
70	N000000682	5-20µm					0,000002549958%		
68	N000000703	5-20µm					0,000009882660%		
68	N000000705	20-180µm					0,000002381136%		
67	N000000756	0.8-5µm			0,000017441181%	0,000011751252%			
67	N000000759	>0.8µm			0,000006466823%	0,000014354905%			
66	N000000793	20-180µm					0,000002445755%		
66	N000000809	5-20µm					0,000026032108%		
84	N000001006	>0.8µm			0,000062312484%	0,000075583896%		0,000028887445%	0,000008584785%
85	N000001017	20-180µm		0,000003004731%				0,000003512533%	0,000046856931%
85	N000001019	20-180µm		0,000000000000%				0,000060044439%	0,000008725951%
85	N000001024	>0.8µm		0,000021693927%	0,000003390557%	0,000002702373%		0,000004861522%	0,000003972097%
85	N000001026	>0.8µm		0,000015065553%	0,000003003850%			0,000003003850%	0,000003003850%
85	N000001029	0.8-5µm		0,000038644238%				0,000002967579%	
85	N000001030	0.8-5µm		0,000041733068%				0,000007999438%	0,000005980182%
85	N000001034	5-20µm		0,000025884379%				0,000014333436%	0,000024543032%
85	N000001040	180-2000µm						0,000004401290%	0,000004317292%
92	N000001299	0.8-5µm	0,000012526918%						
92	N000001301	5-20µm		0,000006584172%					
92	N000001309	>0.8µm	0,000007731768%						
83	N000001374	0.8-5µm			0,000008140644%	0,000009001104%		0,000004044879%	0,000003219805%
83	N000001377	5-20µm			0,000007752460%	0,000014289624%		0,000009961900%	0,000007734404%
82	N000001382	>0.8µm			0,000000000000%	0,000011578650%		0,000009845767%	0,000008331522%
82	N000001384	>0.8µm			0,000003372012%	0,000003886693%		0,000003367624%	0,000007422937%
82	N000001386	0.8-5µm			0,000004413943%	0,000006013370%			
82	N000001390	5-20µm						0,000042583688%	0,000064007097%
82	N000001398	180-2000µm						0,000008526224%	
81	N000001426	20-180µm							0,000002246729%
81	N000001430	5-20µm					0,000002513586%		
81	N000001436	0.8-5µm			0,000005244980%	0,000007422040%			
84	N000001438	0.8-5µm		0,000011935881%	0,000126038462%	0,000076969622%		0,000021638973%	
84	N000001440	5-20µm			0,000010543246%	0,000007259260%		0,000334299132%	0,000039776195%
84	N000001442	20-180µm			0,000006984214%	0,000008372445%		0,000083978421%	0,000020168247%
102	N000001650	0.8-5µm	0,000003514781%						
110	N000001750	0.8-5µm	0,000006920815%						
110	N000001752	0.8-5µm	0,000027946082%						
111	N000001816	0.8-5µm						0,000005647693%	
152	N000002480	>0.8µm			0,000003121599%	0,000004329771%	0,000003676224%		
152	N000002482	>0.8µm			0,000002712505%	0,000002767700%			
152	N000002761	5-20µm			0,000007941840%	0,000004438865%	0,000015437257%		
152	N000002773	>0.8µm				0,000002609195%			
152	N000002775	>0.8µm			0,000003079071%				
152	N000002781	20-180µm					0,000002684583%		
152	N000002783	20-180µm				0,000002197639%			
152	N000002791	0.8-5µm			0,000002164937%	0,000003036921%			
152	N000002797	5-20µm					0,000002874268%		
152	N000002803	>0.8µm			0,000003135760%	0,000003072793%			
153	N000002819	0.8-5µm			0,000018282415%	0,000025230020%	0,000002718281%		
153	N000002823	>0.8µm			0,000002776809%	0,000003393881%			
153	N000002825	>0.8µm				0,000002740357%	0,000001804296%		
136	N000002961	0.8-5µm	0,000002601058%						
139	N000003037	0.8-5µm	0,000002692913%						

			MRM2		
			Bacillariales	Bacillariaceae	Asterionellopsis glacialis
Stations	Samples	Fractions	112166270	112838911	33809363
22	A100000538	20-180µm			0,000013731600144%
22	A100000539	5-20µm			0,000005359750678%
22	A100000695	20-180µm			0,000050311888697%
67	N000000756	0.8-5µm	0,000035231942655%		
67	N000000759	>0.8µm	0,000006152114878%		
84	N000001006	>0.8µm	0,000023573449448%	0,000039367140341%	
85	N000001017	20-180µm		0,000029240453151%	
85	N000001019	20-180µm		0,000008630663705%	
85	N000001024	>0.8µm	0,000003790829455%	0,000002709076057%	
85	N000001026	>0.8µm		0,000004740446689%	
85	N000001029	0.8-5µm	0,000003530061921%	0,000003177051939%	
85	N000001034	5-20µm		0,000007435907852%	
85	N000001040	180-2000µm		0,000005922153710%	
83	N000001374	0.8-5µm	0,000006344924679%		
82	N000001382	>0.8µm	0,000004694043011%	0,000003439200348%	
82	N000001384	>0.8µm	0,000004882452423%	0,000002924637113%	
82	N000001386	0.8-5µm	0,000007151923195%	0,000003139291427%	
81	N000001424	180-2000µm	0,000004235104875%	0,000002419462115%	
81	N000001428	20-180µm	0,000004647508773%		
81	N000001430	5-20µm	0,000002335025689%		
81	N000001436	0.8-5µm	0,000010800006622%		
84	N000001438	0.8-5µm	0,000068296407912%	0,000013020923001%	
84	N000001440	5-20µm	0,000005500802905%	0,000147959396733%	
84	N000001442	20-180µm	0,000006390065047%	0,000060157054831%	
111	N000001823	20-180µm		0,000002304428655%	
152	N000002480	>0.8µm	0,000002576393270%		
152	N000002773	>0.8µm	0,000002208771149%		
152	N000002775	>0.8µm	0,000003548760272%		
152	N000002797	5-20µm	0,000005426183566%		
153	N000002807	180-2000µm	0,000003488447763%		
153	N000002819	0.8-5µm	0,000015784485731%		
153	N000002823	>0.8µm	0,000002458343028%		
153	N000002825	>0.8µm	0,000002413695802%		

			MRP1						
			Bacillariophyceae	Bacillariophyceae	Bacillariophyceae	Bacillariophyceae	Bacillariophyceae	Bacillariophyceae	Bacillariophyceae
Stations	Samples	Fractions	27613804	27666306	27670842	27686307	29402207	80515699	80699714
25	A100000393	0.8-5µm	0,00002017%	0,00003042%	0,00004332%	0,00004662%	0,00003640%	0,00000733%	
25	A100000394	5-20µm	0,00000270%		0,00000453%	0,00000360%			
25	A100000396	0.8-5µm			0,00000695%	0,00001095%	0,00000538%		
25	A100000398	5-20µm		0,00001037%		0,00000292%			
22	A100000534	0.8-5µm	0,00000250%	0,00000783%	0,00000403%	0,00000800%	0,00000411%		
23	A100000551	0.8-5µm	0,00001885%	0,00001864%	0,00002198%	0,00002733%	0,00001559%		
23	A100000554	180-2000µm		0,00000263%		0,00000253%			
23	A100000559	0.8-5µm	0,00001240%	0,00001204%	0,00001744%	0,00002374%	0,00002062%	0,00000331%	
18	A100000595	0.8-5µm	0,00001059%	0,00003713%	0,00001383%	0,00001311%	0,00001734%		
18	A100000596	20-180µm		0,00000631%					
18	A100000598	5-20µm	0,00001276%	0,00000557%	0,00000734%	0,00001154%	0,00001564%		
18	A100000604	0.8-5µm	0,00000958%	0,00000506%	0,00000448%	0,00000481%			
22	A100000696	5-20µm	0,00001426%	0,00001391%	0,00000468%	0,00000415%	0,00002249%		
22	A100000698	0.8-5µm	0,00001201%	0,00001875%	0,00002254%	0,00002308%	0,00002254%	0,00000259%	
20	A100000761	5-20µm	0,00000341%	0,00001119%	0,00000484%	0,00001107%	0,00000953%		
7	A200000123	0.8-5µm	0,00015967%	0,00029717%	0,00008780%	0,00029789%	0,00027410%		
7	A200000139	20-180µm	0,00000381%	0,00000676%	0,00000482%	0,00001231%	0,00001281%		
7	A200000150	5-20µm			0,00001215%	0,00000886%	0,00002621%		
7	A200000158	0.8-5µm	0,00005380%	0,00004711%	0,00005460%	0,00009033%	0,00003703%		
26	E400007200	0.8-20µm	0,00000000%	0,00002394%	0,00001500%	0,00003420%	0,00001352%		
39	N000000006	0.8-5µm	0,00000490%	0,00000393%	0,00001146%	0,00000703%	0,00000968%	0,00000263%	
39	N000000007	0.8-5µm				0,00004010%			
39	N000000008	5-20µm		0,00000567%	0,00000755%	0,00000723%			
38	N000000030	0.8-5µm						0,00000277%	0,00000274%
38	N000000046	180-2000µm						0,00000283%	
38	N000000047	0.8-5µm	0,00000269%			0,00000254%		0,00000756%	
40	N000000063	>3µm	0,00000710%	0,00001598%	0,00001315%	0,00001748%	0,00001077%	0,00000439%	
40	N000000064	>3µm	0,00000282%	0,00000689%	0,00000291%	0,00001697%	0,00000872%		
40	N000000065	>3µm	0,00000459%	0,00000922%	0,00001296%	0,00001488%	0,00000885%		
41	N000000071	0.8-5µm		0,00001739%					
41	N000000073	5-20µm			0,00000986%				
48	N000000113	0.8-20µm	0,00000270%	0,00000286%	0,00000596%	0,00000493%			
51	N000000185	5-20µm		0,00000692%			0,00001189%		
51	N000000216	0.8-5µm							
51	N000000231	0.8-5µm	0,00000613%	0,00001338%	0,00000722%	0,00001083%	0,00000428%		
46	N000000267	0.8-5µm					0,00000387%		
46	N000000269	20-180µm				0,00000795%			
47	N000000278	0.8-20µm	0,00001513%	0,00000927%		0,00001657%	0,00000687%		
36	N000000312	20-180µm				0,00000000%			
36	N000000316	0.8-5µm			0,00000650%	0,00000323%			
36	N000000317	0.8-5µm						0,00000362%	
36	N000000319	5-20µm							
64	N000000522	0.8-5µm							
64	N000000526	20-180µm							
64	N000000531	5-20µm						0,00002971%	0,00000515%
64	N000000535	20-180µm							
52	N000000570	5-20µm			0,00001264%	0,00000299%	0,00000569%	0,00000778%	
52	N000000589	20-180µm	0,00000244%						
52	N000000593	20-180µm							
52	N000000600	0.8-5µm						0,00000506%	
70	N000000662	20-180µm							
70	N000000674	>0.8µm		0,00000244%				0,00002239%	0,00001184%
70	N000000678	0.8-5µm						0,00008112%	0,00004044%
70	N000000682	5-20µm			0,00000510%			0,00000619%	
68	N000000701	5-20µm							
68	N000000703	5-20µm						0,00000838%	0,00000264%
68	N000000705	20-180µm							
68	N000000720	>0.8µm		0,00000261%				0,00000449%	
68	N000000722	0.8-5µm						0,00000510%	0,00000424%
68	N000000728	0.8-5µm		0,00000235%				0,00000766%	0,00000669%
67	N000000759	>0.8µm							
66	N000000790	180-2000µm							
66	N000000793	20-180µm							
66	N000000807	0.8-5µm							
66	N000000809	5-20µm							
72	N000000836	5-20µm			0,00000513%		0,00001301%		0,00000298%
64	N000000927	5-20µm							
65	N000000933	0.8-5µm					0,00000310%		
65	N000000939	5-20µm							
65	N000000963	5-20µm							
65	N000000968	20-180µm							
82	N000001382	>0.8µm						0,00001039%	0,00000344%
82	N000001384	>0.8µm						0,00000292%	0,00000292%
82	N000001386	0.8-5µm						0,00000462%	
81	N000001424	180-2000µm							
81	N000001426	20-180µm							
81	N000001428	20-180µm							
81	N000001430	5-20µm			0,00000212%		0,00000212%	0,00000305%	
81	N000001436	0.8-5µm							
80	N000001478	>0.8µm	0,00000310%				0,00000457%	0,00000281%	
80	N000001491	0.8-5µm					0,00000356%	0,00001058%	
80	N000001503	180-2000µm						0,00000252%	
98	N000001578	5-20µm							
98	N000001580	5-20µm							
98	N000001582	20-180µm							
100	N000001604	>0.8µm						0,00000306%	0,00000368%
100	N000001606	>0.8µm						0,00000306%	
100	N000001608	0.8-5µm						0,00000466%	0,00000230%
100	N000001610	0.8-5µm						0,00000357%	
100	N000001612	5-20µm			0,00000258%				
100	N000001614	5-20µm						0,00000529%	
100	N000001616	20-180µm							
100	N000001618	20-180µm							
100	N000001643	>3µm							

			MRP1						
			Bacillariophyceae	Bacillariophyceae	Bacillariophyceae	Bacillariophyceae	Bacillariophyceae	Bacillariophyceae	Bacillariophyceae
Stations	Samples	Fractions	27613804	27666306	27670842	27686307	29402207	80515699	80699714
102	N000001646	>0.8µm		0,00000249%	0,00000243%		0,00000226%	0,00000423%	0,00000307%
102	N000001648	>0.8µm		0,00000312%	0,00000198%			0,00000513%	
102	N000001650	0.8-5µm			0,00000294%		0,00000238%	0,00001041%	
102	N000001652	0.8-5µm						0,00000000%	
102	N000001654	5-20µm		0,00000402%				0,00000804%	
102	N000001656	5-20µm			0,00000550%	0,00000322%	0,00000725%	0,00001484%	0,00000263%
102	N000001658	20-180µm							
102	N000001660	20-180µm							
102	N000001662	180-2000µm					0,00000229%	0,00000323%	
109	N000001727	>0.8µm				0,00000356%		0,00000000%	
109	N000001728	>0.8µm			0,00000260%			0,00000550%	0,00000260%
109	N000001730	0.8-5µm				0,00000295%		0,00000270%	
109	N000001732	0.8-5µm		0,00000305%				0,00000458%	
109	N000001738	20-180µm							
109	N000001740	20-180µm							
110	N000001746	>0.8µm						0,00000739%	
110	N000001748	>0.8µm					0,00000639%	0,00000835%	0,00000621%
110	N000001750	0.8-5µm					0,00000756%	0,00002032%	0,00000504%
110	N000001752	0.8-5µm				0,00000296%	0,00000502%	0,00006481%	0,00003735%
110	N000001754	5-20µm			0,00000630%		0,00002046%	0,00000727%	0,00000418%
110	N000001758	20-180µm							
111	N000001810	>0.8µm		0,00000244%	0,00000278%		0,00000518%	0,00000410%	0,00000244%
111	N000001812	0.8-5µm						0,00002554%	
111	N000001814	>0.8µm			0,00000282%				
111	N000001816	0.8-5µm							
111	N000001823	20-180µm							
111	N000001824	20-180µm							
122	N000001937	>0.8µm		0,00000478%			0,00000317%	0,00001058%	0,00000427%
122	N000001938	0.8-5µm		0,00000255%			0,00000257%	0,00002625%	0,00001510%
122	N000001941	20-180µm							
122	N000001942	180-2000µm							
122	N000001943	5-20µm							
123	N000001972	>0.8µm					0,00000193%	0,00000726%	0,00000295%
123	N000001992	0.8-5µm		0,00000235%				0,00000711%	0,00000405%
123	N000001994	5-20µm							
123	N000001996	20-180µm							
125	N000002017	>0.8µm		0,00000377%			0,00000198%	0,00000468%	0,00000258%
125	N000002019	0.8-5µm			0,00000189%			0,00000555%	
125	N000002021	5-20µm							
125	N000002024	20-180µm							
124	N000002036	>0.8µm		0,00000398%			0,00000446%	0,00000363%	
124	N000002037	0.8-5µm		0,00000220%			0,00000567%	0,00000631%	
124	N000002041	20-180µm							
135	N000002179	0.8-5µm			0,00000425%	0,00000309%	0,00000246%		
128	N000002285	>0.8µm			0,00000290%	0,00000194%	0,00000355%	0,00001303%	0,00000794%
128	N000002287	>0.8µm					0,00000196%	0,00000625%	
128	N000002289	0.8-5µm	0,00000226%	0,00000512%	0,00000226%		0,00000558%	0,00002334%	0,00000391%
128	N000002291	0.8-5µm		0,00000243%			0,00000243%	0,00000955%	0,00000417%
128	N000002293	5-20µm		0,00000664%				0,00000730%	0,00000581%
128	N000002295	5-20µm					0,00000393%	0,00000658%	
131	N000002348	>0.8µm	0,00000394%			0,00000981%	0,00001064%		
131	N000002350	>0.8µm						0,00000622%	
131	N000002352	0.8-5µm			0,00000591%		0,00000432%		
131	N000002354	0.8-5µm			0,00000490%				
132	N000002404	20-180µm				0,00000256%			
132	N000002412	5-20µm					0,00000414%		
132	N000002414	5-20µm							
132	N000002416	0.8-5µm	0,00001232%	0,00004893%	0,00002942%	0,00006334%	0,00006714%		
132	N000002418	0.8-5µm			0,00000783%				
132	N000002420	>0.8µm	0,00001297%	0,00001499%	0,00002688%	0,00003864%	0,00003070%		
132	N000002422	>0.8µm						0,00000370%	
135	N000002464	>0.8µm			0,00000319%		0,00000470%	0,00000719%	0,00000325%
152	N000002761	5-20µm					0,00000261%		
137	N000002923	>0.8µm				0,00000406%	0,00000229%		
137	N000002925	0.8-5µm							0,00000228%
137	N000002934	>0.8µm							
137	N000002940	20-180µm							
136	N000002959	>0.8µm	0,00000418%	0,00000633%	0,00000370%	0,00000703%	0,00000235%	0,00000376%	
136	N000002961	0.8-5µm		0,00000438%	0,00000288%	0,00000520%	0,00000261%	0,00000478%	
136	N000002963	5-20µm		0,00000287%	0,00000451%	0,00000287%	0,00000756%	0,00000287%	
136	N000002965	20-180µm							
138	N000003003	5-20µm	0,00000436%	0,00001072%	0,00001163%	0,00002854%	0,00001103%		
138	N000003011	0.8-5µm						0,00000252%	
139	N000003037	0.8-5µm		0,00000267%		0,00000338%			
139	N000003039	5-20µm							
142	N000003081	>0.8µm							
142	N000003087	20-180µm							
142	N000003102	5-20µm							
142	N000003104	0.8-5µm							
144	N000003175	>0.8µm						0,00000306%	
9	X000000954	0.8-5µm		0,00003514%	0,00004550%	0,00003514%	0,00006914%		
9	X000001006	180-2000µm		0,00001080%			0,00000711%		
9	X000001040	0.8-5µm		0,00002752%	0,00003179%				
9	X000001043	5-20µm	0,00001797%	0,00000593%		0,00000599%	0,00000587%		
11	X000001286	5-20µm	0,00001942%	0,00000518%	0,00001933%	0,00002068%	0,00000681%		
11	X000001288	0.8-5µm	0,00004403%	0,00005101%	0,00003677%	0,00005415%	0,00004131%		

			MRP1						
			Bacillariophyceae	Bacillariophyceae	Bacillariophyceae	Bacillariophyceae	Bacillariophyceae	Bacillariophyceae	Bacillariophyceae
Stations	Samples	Fractions	80850470	81079621	82966970	82997029	89189946	96428693	97079723
25	A100000393	0.8-5µm					0,00000593%		
25	A100000394	5-20µm							
25	A100000396	0.8-5µm					0,00000495%		
25	A100000398	5-20µm					0,00000475%		
22	A100000534	0.8-5µm					0,00000590%		
23	A100000551	0.8-5µm					0,00001407%		
23	A100000554	180-2000µm							
23	A100000559	0.8-5µm					0,00000845%		
18	A100000595	0.8-5µm					0,00001127%		
18	A100000596	20-180µm							
18	A100000598	5-20µm					0,00000294%		
18	A100000604	0.8-5µm							
22	A100000696	5-20µm					0,00000909%		
22	A100000698	0.8-5µm					0,00001584%		
20	A100000761	5-20µm							
7	A200000123	0.8-5µm					0,00009765%		
7	A200000139	20-180µm							
7	A200000150	5-20µm							
7	A200000158	0.8-5µm					0,00003819%		
26	E400007200	0.8-20µm					0,00001284%		
39	N000000006	0.8-5µm					0,00000730%		
39	N000000007	0.8-5µm							
39	N000000008	5-20µm							
38	N000000030	0.8-5µm	0,00000336%						
38	N000000046	180-2000µm							
38	N000000047	0.8-5µm	0,00000264%	0,00000294%					
40	N000000063	>3µm					0,00001535%		
40	N000000064	>3µm					0,00000282%		
40	N000000065	>3µm		0,00000240%			0,00000733%		
41	N000000071	0.8-5µm							
41	N000000073	5-20µm							
48	N000000113	0.8-20µm					0,00000309%		
51	N000000185	5-20µm							
51	N000000216	0.8-5µm		0,00000836%					
51	N000000231	0.8-5µm					0,00000751%		
46	N000000267	0.8-5µm							
46	N000000269	20-180µm							
47	N000000278	0.8-20µm					0,00000483%		
36	N000000312	20-180µm							
36	N000000316	0.8-5µm							
36	N000000317	0.8-5µm					0,00000255%		
36	N000000319	5-20µm							
64	N000000522	0.8-5µm							
64	N000000526	20-180µm				0,00000310%		0,00000522%	
64	N000000531	5-20µm		0,00000568%					
64	N000000535	20-180µm			0,00002841%	0,00001015%		0,00000445%	
52	N000000570	5-20µm	0,00000299%		0,00000305%	0,00000498%	0,00000529%	0,00000830%	0,00001057%
52	N000000589	20-180µm			0,00000426%	0,00000429%		0,00000268%	0,00000421%
52	N000000593	20-180µm			0,00002226%	0,00002256%		0,00001136%	0,00000656%
52	N000000600	0.8-5µm							
70	N000000662	20-180µm				0,00000206%		0,00000206%	
70	N000000674	>0.8µm	0,00000443%	0,00001649%					
70	N000000678	0.8-5µm	0,00003029%	0,00005498%					
70	N000000682	5-20µm	0,00000274%	0,00000535%		0,00000271%			
68	N000000701	5-20µm							
68	N000000703	5-20µm							
68	N000000705	20-180µm			0,00000238%	0,00001126%		0,00000814%	0,00000305%
68	N000000720	>0.8µm	0,00000383%						
68	N000000722	0.8-5µm							
68	N000000728	0.8-5µm							
67	N000000759	>0.8µm							
66	N000000790	180-2000µm				0,00000256%			
66	N000000793	20-180µm			0,00001466%	0,00019992%		0,00009800%	0,00005069%
66	N000000807	0.8-5µm				0,00000368%		0,00000384%	
66	N000000809	5-20µm				0,00001922%		0,00003364%	0,00000469%
72	N000000836	5-20µm		0,00000519%			0,00000298%		
64	N000000927	5-20µm							
65	N000000933	0.8-5µm							
65	N000000939	5-20µm		0,00000420%					
65	N000000963	5-20µm			0,00002239%	0,00002142%			
65	N000000968	20-180µm			0,00000843%	0,00000859%		0,00000377%	0,00000721%
82	N000001382	>0.8µm	0,00000572%	0,00000344%			0,00000344%		
82	N000001384	>0.8µm							
82	N000001386	0.8-5µm	0,00000314%						
81	N000001424	180-2000µm			0,00000242%			0,00000242%	
81	N000001426	20-180µm			0,00000337%	0,00003305%		0,00000975%	0,00000639%
81	N000001428	20-180µm				0,00000913%			
81	N000001430	5-20µm	0,00000455%		0,00000212%	0,00001446%		0,00000372%	0,00000337%
81	N000001436	0.8-5µm		0,00000268%			0,00000268%		
80	N000001478	>0.8µm							
80	N000001491	0.8-5µm	0,00000475%						
80	N000001503	180-2000µm							
98	N000001578	5-20µm				0,00000309%			
98	N000001580	5-20µm							
98	N000001582	20-180µm			0,00000339%				0,00000232%
100	N000001604	>0.8µm	0,00000233%		0,00000444%	0,00000233%			
100	N000001606	>0.8µm			0,00002193%				0,00000595%
100	N000001608	0.8-5µm	0,00000464%	0,00000224%					
100	N000001610	0.8-5µm		0,00000896%					
100	N000001612	5-20µm			0,00000506%	0,00000720%			0,00000264%
100	N000001614	5-20µm			0,00000000%	0,00003668%		0,00000529%	
100	N000001616	20-180µm			0,00006019%	0,00006359%		0,00002704%	0,00001654%
100	N000001618	20-180µm			0,00012016%	0,00003070%		0,00004126%	
100	N000001643	>3µm			0,00014060%	0,00000000%		0,00011902%	

			MRP1						
			Bacillariophyceae	Bacillariophyceae	Bacillariophyceae	Bacillariophyceae	Bacillariophyceae	Bacillariophyceae	Bacillariophyceae
Stations	Samples	Fractions	80850470	81079621	82966970	82997029	89189946	96428693	97079723
102	N000001646	>0.8µm	0,00000341%	0,00000898%	0,00000228%	0,00000271%	0,00000226%		0,00000336%
102	N000001648	>0.8µm		0,00000643%	0,00000247%	0,00000304%			
102	N000001650	0.8-5µm	0,00000238%	0,00002862%	0,00000000%	0,00000000%	0,00000758%		
102	N000001652	0.8-5µm	0,00000276%	0,00001654%	0,00000226%	0,00000000%	0,00000226%		
102	N000001654	5-20µm	0,00000000%	0,00000837%	0,00000597%	0,00002336%			
102	N000001656	5-20µm	0,00000388%	0,00006351%	0,00008097%	0,00006557%		0,00002953%	0,00001584%
102	N000001658	20-180µm			0,00001874%	0,00000797%		0,00000545%	0,00000642%
102	N000001660	20-180µm			0,00002139%	0,00001920%		0,00001919%	
102	N000001662	180-2000µm	0,00000229%						
109	N000001727	>0.8µm	0,00000000%						
109	N000001728	>0.8µm	0,00000272%						
109	N000001730	0.8-5µm	0,00000000%						
109	N000001732	0.8-5µm	0,00000726%	0,00000305%					
109	N000001738	20-180µm			0,00000349%	0,00000234%			0,00000000%
109	N000001740	20-180µm			0,00000861%	0,00000465%			0,00000242%
110	N000001746	>0.8µm							
110	N000001748	>0.8µm	0,00000627%	0,00000781%					
110	N000001750	0.8-5µm	0,00000690%						
110	N000001752	0.8-5µm	0,00003368%	0,00003034%					
110	N000001754	5-20µm	0,00000418%			0,00000418%		0,00000879%	
110	N000001758	20-180µm			0,00000342%				
111	N000001810	>0.8µm	0,00000460%		0,00000340%	0,00000332%	0,00000487%		
111	N000001812	0.8-5µm	0,00001700%	0,00000403%					
111	N000001814	>0.8µm	0,00000343%	0,00001330%	0,00000282%	0,00000290%			
111	N000001816	0.8-5µm	0,00000386%	0,00000982%	0,00000386%				
111	N000001823	20-180µm			0,00023793%	0,00013822%		0,00002855%	0,00003362%
111	N000001824	20-180µm			0,00001882%	0,00001266%		0,00000348%	0,00000462%
122	N000001937	>0.8µm	0,00000688%		0,00000549%	0,00000221%	0,00000372%		
122	N000001938	0.8-5µm	0,00000457%						
122	N000001941	20-180µm			0,00002725%			0,00001344%	
122	N000001942	180-2000µm							
122	N000001943	5-20µm			0,00001921%	0,00000938%		0,00001269%	0,00000573%
123	N000001972	>0.8µm							
123	N000001992	0.8-5µm	0,00000235%						
123	N000001994	5-20µm	0,00000644%				0,00000263%		
123	N000001996	20-180µm	0,00000268%		0,00000488%	0,00000203%	0,00000000%		
125	N000002017	>0.8µm					0,00000232%		
125	N000002019	0.8-5µm	0,00000189%						
125	N000002021	5-20µm				0,00000235%		0,00000296%	
125	N000002024	20-180µm							
124	N000002036	>0.8µm					0,00000433%		
124	N000002037	0.8-5µm	0,00000213%	0,00000326%					
124	N000002041	20-180µm			0,00000760%				
135	N000002179	0.8-5µm							
128	N000002285	>0.8µm	0,00000505%	0,00000194%			0,00000322%		
128	N000002287	>0.8µm					0,00000196%		
128	N000002289	0.8-5µm	0,00001051%	0,00000567%			0,00000226%		
128	N000002291	0.8-5µm	0,00000523%	0,00000243%					
128	N000002293	5-20µm							
128	N000002295	5-20µm	0,00000606%						
131	N000002348	>0.8µm							
131	N000002350	>0.8µm		0,00000536%					
131	N000002352	0.8-5µm		0,00000407%					
131	N000002354	0.8-5µm	0,00000306%	0,00003604%			0,00000315%		
132	N000002404	20-180µm							
132	N000002412	5-20µm							
132	N000002414	5-20µm					0,00000347%		
132	N000002416	0.8-5µm					0,00002912%		
132	N000002418	0.8-5µm		0,00007014%			0,00000399%		
132	N000002420	>0.8µm					0,00001153%		
132	N000002422	>0.8µm		0,00002275%					
135	N000002464	>0.8µm			0,00001181%	0,00000345%	0,00000322%	0,00000328%	
152	N000002761	5-20µm							
137	N000002923	>0.8µm	0,00000229%						
137	N000002925	0.8-5µm							
137	N000002934	>0.8µm		0,00000332%					
137	N000002940	20-180µm			0,00000288%	0,00000209%			0,00000299%
136	N000002959	>0.8µm					0,00000235%		
136	N000002961	0.8-5µm		0,00000260%			0,00000260%		
136	N000002963	5-20µm	0,00000302%				0,00000287%		
136	N000002965	20-180µm			0,00000262%				
138	N000003003	5-20µm					0,00000862%		
138	N000003011	0.8-5µm	0,00000252%	0,00001181%			0,00000255%		
139	N000003037	0.8-5µm		0,00000267%			0,00000269%		
139	N000003039	5-20µm				0,00000311%			
142	N000003081	>0.8µm				0,00000322%			
142	N000003087	20-180µm			0,00000399%				
142	N000003102	5-20µm			0,00001895%	0,00000746%		0,00000629%	0,00000274%
142	N000003104	0.8-5µm							
144	N000003175	>0.8µm		0,00000306%					
9	X000000954	0.8-5µm							
9	X000001006	180-2000µm					0,00002123%		
9	X000001040	0.8-5µm					0,00000913%		
9	X000001043	5-20µm					0,00001638%		
11	X000001286	5-20µm							
11	X000001288	0.8-5µm					0,00002290%		

			MRP2			
			Bacillariales	Bacillariales	Bacillariophyceae	Bacillariophyta
Stations	Samples	Fractions	97117869	83523918	53436960	44070075
64	N000000526	20-180µm	0,000003101%			
52	N000000589	20-180µm	0,000002494%			
52	N000000593	20-180µm	0,000003454%	0,000003006%		
52	N000000600	0.8-5µm		0,000007547%		
67	N000000742	5-20µm			0,000013109%	
67	N000000746	20-180µm			0,000004856%	
67	N000000756	0.8-5µm			0,000052354%	
67	N000000759	>0.8µm			0,000052308%	
66	N000000793	20-180µm	0,000036128%			
65	N000000968	20-180µm	0,000002425%			0,000002%
84	N000001006	>0.8µm			0,000064720%	
85	N000001029	0.8-5µm			0,000003408%	
85	N000001030	0.8-5µm			0,000003049%	
85	N000001034	5-20µm			0,000007664%	
83	N000001374	0.8-5µm			0,000004334%	
82	N000001382	>0.8µm			0,000003456%	
82	N000001384	>0.8µm			0,000004370%	
82	N000001386	0.8-5µm			0,000005222%	
82	N000001390	5-20µm			0,000018415%	
81	N000001424	180-2000µm			0,000005060%	
81	N000001426	20-180µm	0,000002891%		0,000003277%	
81	N000001428	20-180µm			0,000010414%	
81	N000001430	5-20µm			0,000003755%	
81	N000001436	0.8-5µm			0,000019989%	0,000003%
84	N000001438	0.8-5µm			0,000100045%	
84	N000001440	5-20µm			0,000016832%	
84	N000001442	20-180µm			0,000006820%	
80	N000001489	5-20µm	0,000015560%			
100	N000001604	>0.8µm	0,000004362%	0,000014722%		
100	N000001606	>0.8µm		0,000009021%		
100	N000001610	0.8-5µm		0,000004437%		
100	N000001612	5-20µm	0,000002776%	0,000008712%		
100	N000001614	5-20µm		0,000016303%		
100	N000001616	20-180µm	0,000004259%	0,000008634%		
100	N000001618	20-180µm	0,000011340%			
102	N000001646	>0.8µm	0,000003237%	0,000003484%		
102	N000001656	5-20µm	0,000007833%			
102	N000001658	20-180µm	0,000004006%			
109	N000001740	20-180µm	0,000006873%			
110	N000001758	20-180µm	0,000002632%			
111	N000001810	>0.8µm	0,000002780%	0,000002946%		
111	N000001814	>0.8µm		0,000002817%		
111	N000001823	20-180µm	0,000014393%			
111	N000001824	20-180µm	0,000004279%			
122	N000001937	>0.8µm		0,000019661%		
122	N000001941	20-180µm		0,000015363%		
122	N000001943	5-20µm		0,000065036%		
123	N000001972	>0.8µm		0,000002043%		
123	N000001992	0.8-5µm		0,000006443%		
123	N000001994	5-20µm		0,000015791%		
123	N000001996	20-180µm	0,000002022%	0,000005189%		
125	N000002017	>0.8µm		0,000005334%		
125	N000002019	0.8-5µm		0,000002805%		
125	N000002021	5-20µm	0,000002398%	0,000002626%		
124	N000002036	>0.8µm		0,000004877%		
124	N000002037	0.8-5µm		0,000002816%		
128	N000002285	>0.8µm		0,000005936%		
128	N000002287	>0.8µm		0,000002746%		0,000002%
128	N000002289	0.8-5µm		0,000002263%		
128	N000002291	0.8-5µm		0,000003490%		
128	N000002293	5-20µm		0,000017084%		
128	N000002295	5-20µm		0,000004225%		
135	N000002464	>0.8µm		0,000004680%		
152	N000002482	>0.8µm			0,000003174%	0,000004%
152	N000002761	5-20µm	0,000005221%			
152	N000002773	>0.8µm			0,000004119%	0,000004%
152	N000002775	>0.8µm			0,000002857%	0,000004%
152	N000002789	0.8-5µm			0,000004874%	0,000005%
152	N000002791	0.8-5µm			0,000002290%	0,000003%
152	N000002803	>0.8µm			0,000003632%	0,000019%
153	N000002807	180-2000µm			0,000003242%	0,000003%
153	N000002819	0.8-5µm			0,000092550%	0,000014%
153	N000002823	>0.8µm			0,000005554%	0,000008%
153	N000002825	>0.8µm			0,000002652%	0,000003%
137	N000002923	>0.8µm		0,000002574%		
137	N000002938	5-20µm		0,000002749%		
142	N000003102	5-20µm	0,000004428%			
4	X000000408	0.8-5µm			0,000003358%	0,000005%

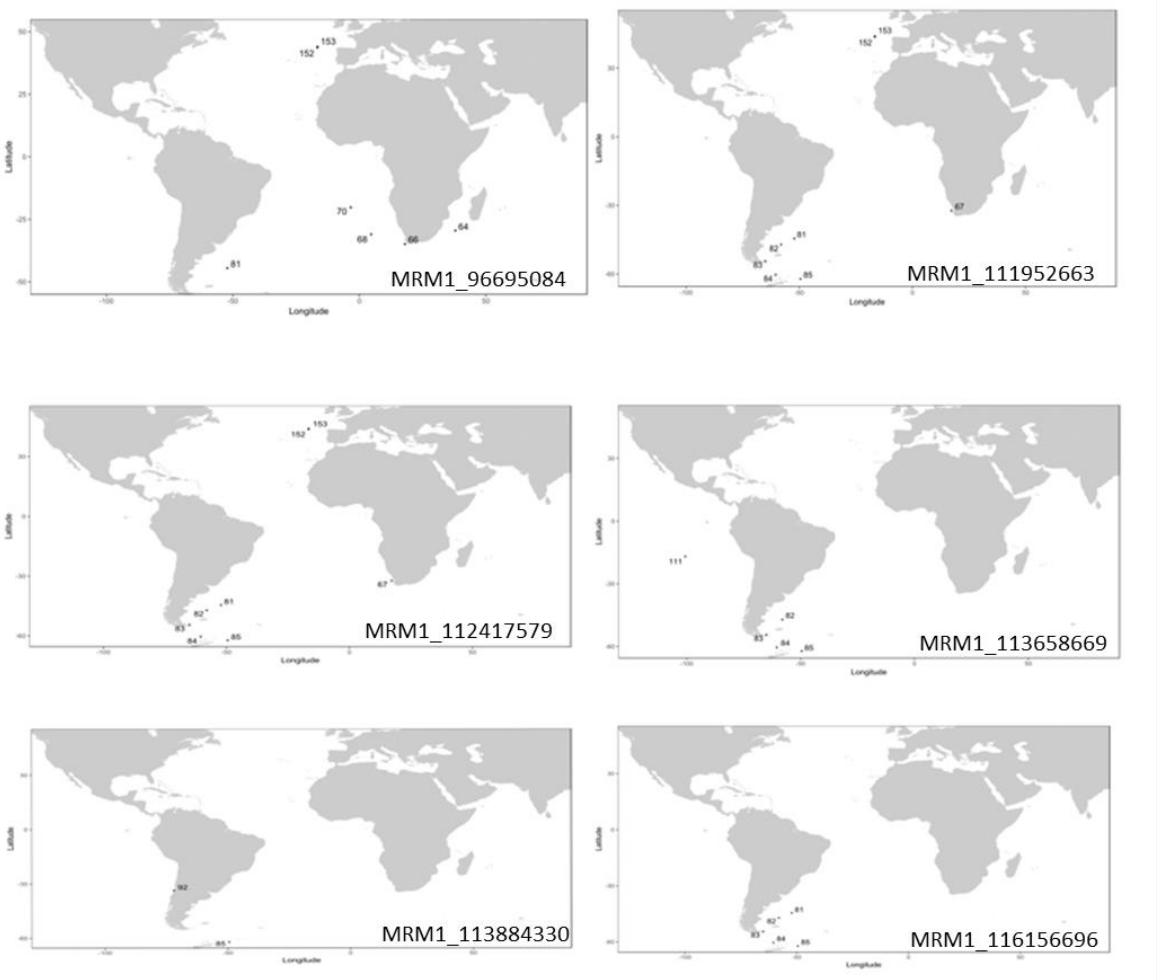
			MRP3					
			Bacillariophyceae	Bacillariophyceae	Bacillariophyceae	Bacillariaceae	Fragilariopsis	Pseudo-nitzschia
Stations	Samples	Fractions	90789615	110585672	112090697	112989460	113648004	117088931
38	N000000041	20-180µm	0,00000340%					
64	N000000531	5-20µm	0,00002923%					
67	N000000746	20-180µm			0,00000516%			
67	N000000756	0.8-5µm			0,00002727%			
67	N000000759	>0.8µm			0,00002126%			
84	N000001006	>0.8µm			0,00002619%	0,00001514%	0,00001690%	0,00000333%
85	N000001017	20-180µm				0,00000952%		0,00000248%
85	N000001019	20-180µm						0,00000317%
85	N000001024	>0.8µm						0,00000695%
85	N000001026	>0.8µm			0,00000300%			
85	N000001029	0.8-5µm			0,00000485%			0,00001809%
85	N000001030	0.8-5µm						0,00000305%
85	N000001034	5-20µm					0,00000578%	0,00001193%
85	N000001040	180-2000µm				0,00000280%	0,00000374%	
92	N000001299	0.8-5µm		0,00000290%				
84	N000001362	180-2000µm					0,00000388%	
83	N000001374	0.8-5µm			0,00000410%			
82	N000001382	>0.8µm					0,00000547%	
82	N000001384	>0.8µm		0,00000450%	0,00000700%	0,00000613%	0,00000487%	
82	N000001386	0.8-5µm			0,00000635%		0,00000401%	
82	N000001390	5-20µm		0,00004020%		0,00000561%	0,00000350%	
82	N000001398	180-2000µm			0,00000343%			
81	N000001424	180-2000µm			0,00000383%			
81	N000001426	20-180µm		0,00000395%				
81	N000001428	20-180µm			0,00000554%			
81	N000001436	0.8-5µm			0,00001589%			
84	N000001438	0.8-5µm			0,00009655%	0,00000612%	0,00001729%	
84	N000001440	5-20µm			0,00000438%	0,00003951%	0,00015139%	
84	N000001442	20-180µm			0,00000657%	0,00006062%	0,00001570%	
80	N000001499	20-180µm						
80	N000001503	180-2000µm						
102	N000001648	>0.8µm	0,00000198%					
109	N000001728	>0.8µm	0,00000278%					
122	N000001937	>0.8µm	0,00000221%					
125	N000002019	0.8-5µm	0,00000193%					
152	N000002482	>0.8µm	0,00000259%		0,00000243%			
152	N000002775	>0.8µm			0,00000241%			
152	N000002789	0.8-5µm			0,00000355%			
152	N000002803	>0.8µm			0,00000253%			
153	N000002813	20-180µm			0,00000264%			
153	N000002819	0.8-5µm			0,00003217%			
153	N000002823	>0.8µm			0,00000277%			
153	N000002825	>0.8µm			0,00000344%			
138	N000003011	0.8-5µm	0,00000252%					
139	N000003037	0.8-5µm	0,00000296%					
144	N000003183	5-20µm	0,00000350%					

			SPO11-2				
			Bacillariophyta	Thalassiosiraceae	Bacillariophyta	Bacillariophyta	Bacillariophyta
Stations	Samples	Fractions	109206300	90705914	91223600	111937145	112625650
64	5-20µm	N000000531		0,00006618547%	0,00003242567%		
67	0.8-5µm	N000000756				0,00001529988%	0,00001476393%
67	>0.8µm	N000000759				0,00001309084%	0,00000731684%
84	>0.8µm	N000001006	0,00000917689%			0,00003966822%	0,00002978324%
85	20-180µm	N000001017	0,00001360987%				
85	20-180µm	N000001019	0,00000677420%				
85	>0.8µm	N000001024	0,00000315537%			0,00000279938%	
85	0.8-5µm	N000001029				0,00000285013%	
83	0.8-5µm	N000001374				0,00000390652%	
82	>0.8µm	N000001382	0,00000672310%				
82	>0.8µm	N000001384	0,00002571460%			0,00000292464%	
82	180-2000µm	N000001398	0,00000276527%				
81	0.8-5µm	N000001436				0,00000609559%	0,00000399249%
84	0.8-5µm	N000001438				0,00007809035%	0,00002017833%
84	5-20µm	N000001440	0,00003891719%			0,00001533051%	0,00002139648%
84	20-180µm	N000001442	0,00002287355%			0,00000910843%	0,00000510721%
152	>0.8µm	N000002775				0,00000254304%	
152	0.8-5µm	N000002791					0,00000238971%
152	>0.8µm	N000002803				0,00000459535%	
153	0.8-5µm	N000002819				0,00001233184%	0,00000319240%
153	>0.8µm	N000002823				0,00000349084%	0,00000000000%
153	>0.8µm	N000002825					0,00000180430%
142	5-20µm	N000003102	0,000020858%				

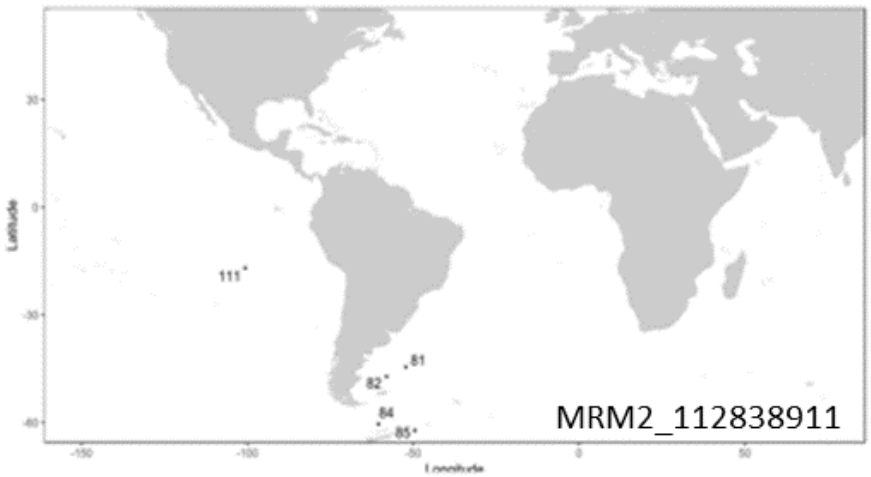
Appendix 2

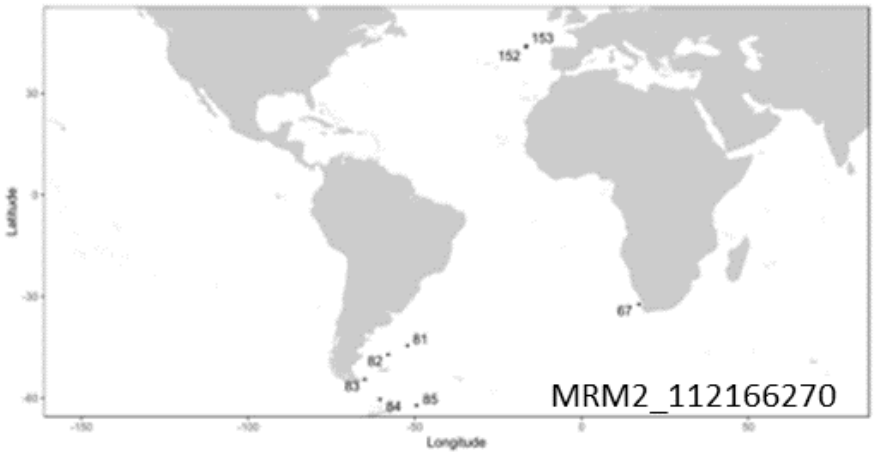
Maps of homologs

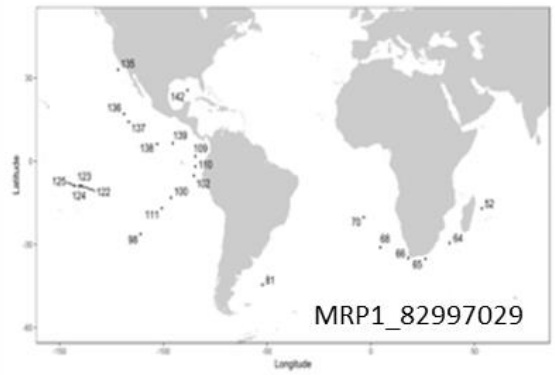
Distribution maps of each homolog of MR genes in TARA stations.

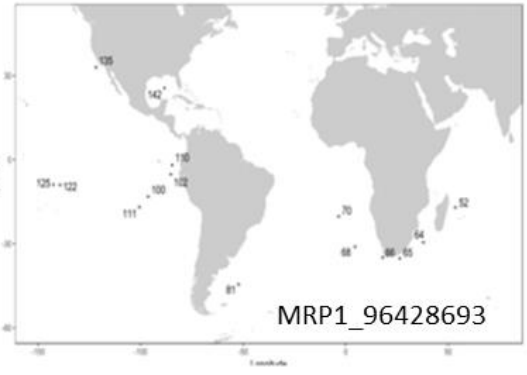


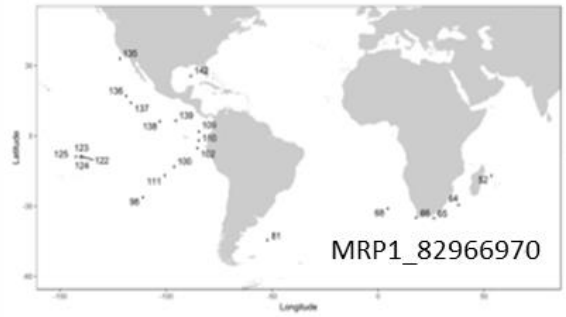

22
Latitude
MRM2_33809363

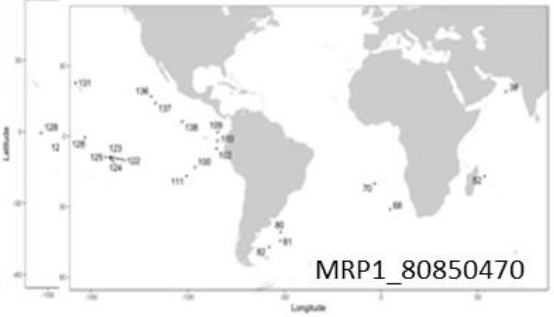

111
81
82
84
85
Latitude
MRM2_112838911

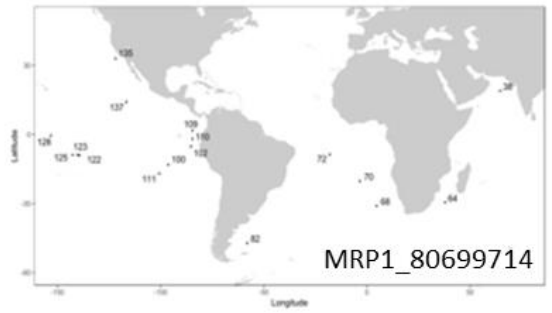

153
152
67
81
82
83
84
85
Latitude
Longitude
MRM2_112166270

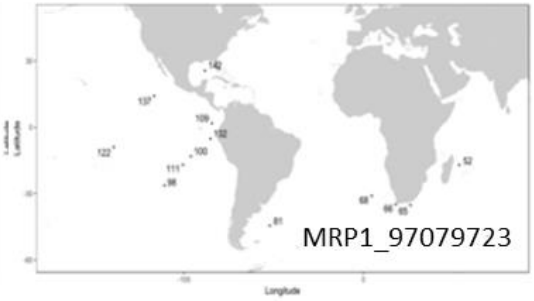

MRP1_82997029

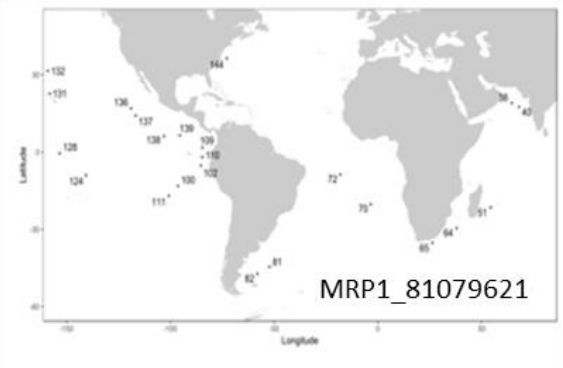

MRP1_96428693

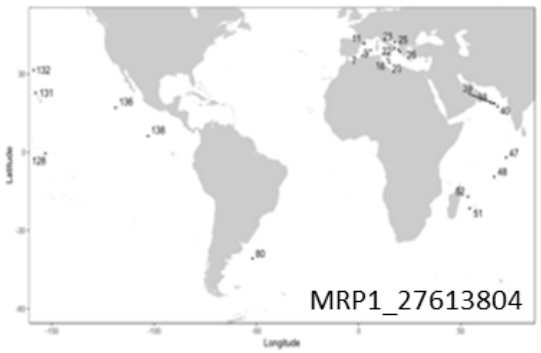

MRP1_82966970

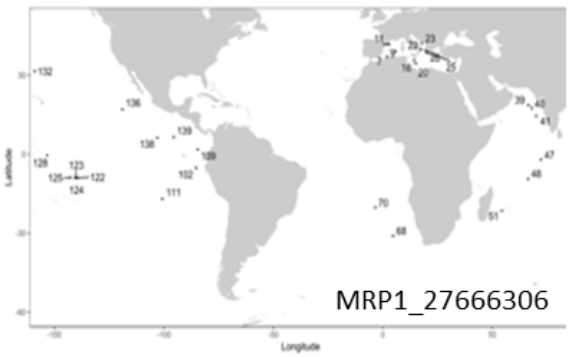

MRP1_80850470

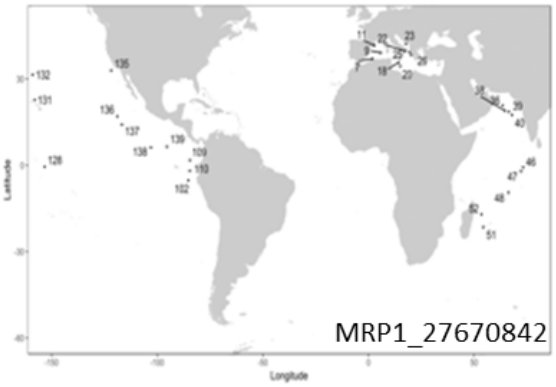

MRP1_80699714

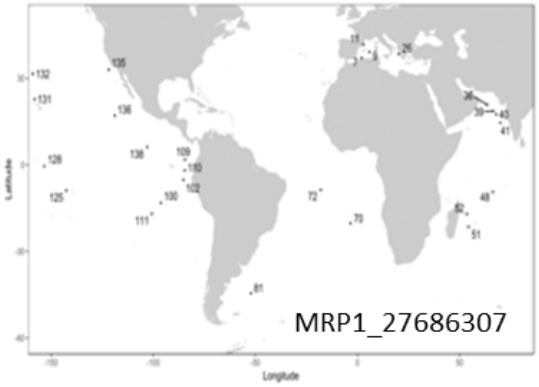

MRP1_97079723

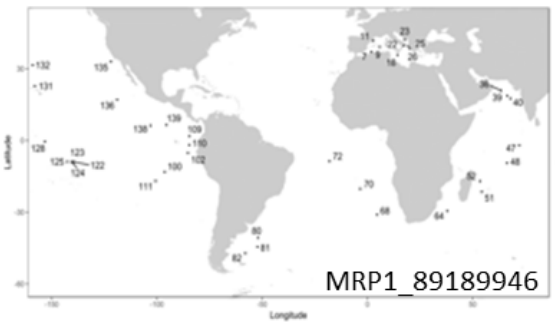

MRP1_81079621

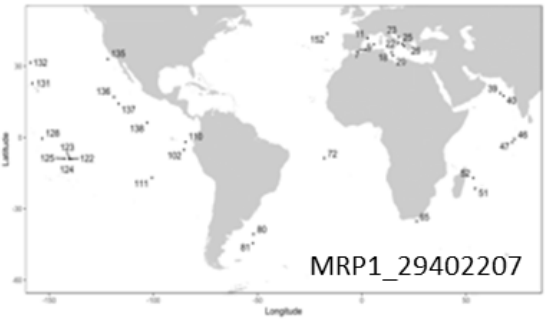

MRP1_27613804
Latitude
Longitude

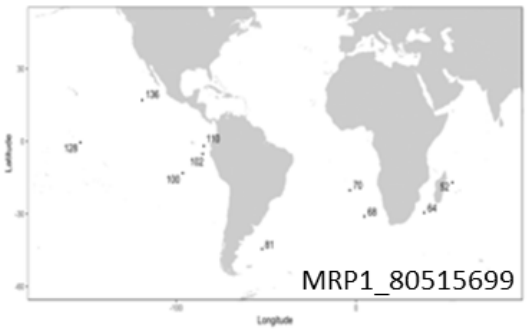

MRP1_27666306
Latitude
Longitude

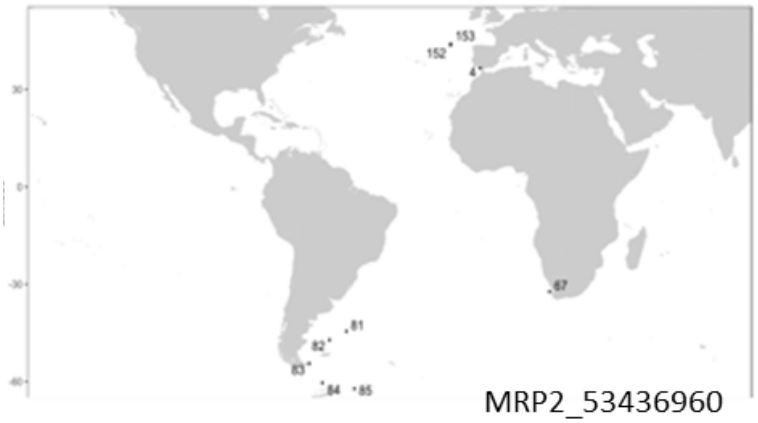

MRP1_27670842
Latitude
Longitude

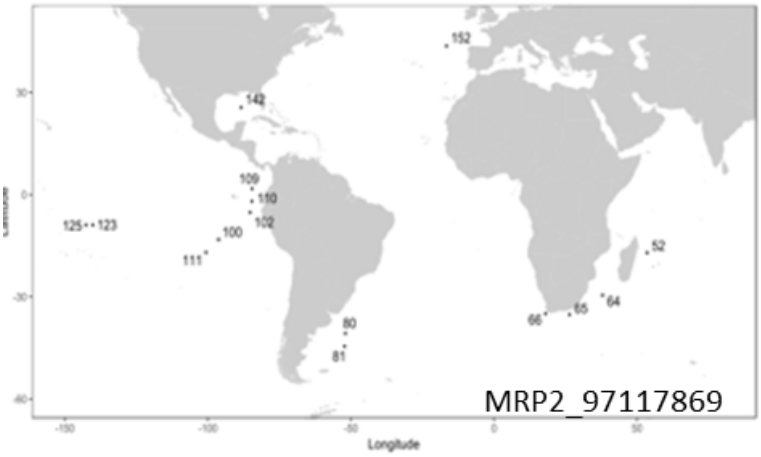

MRP1_27686307
Latitude
Longitude

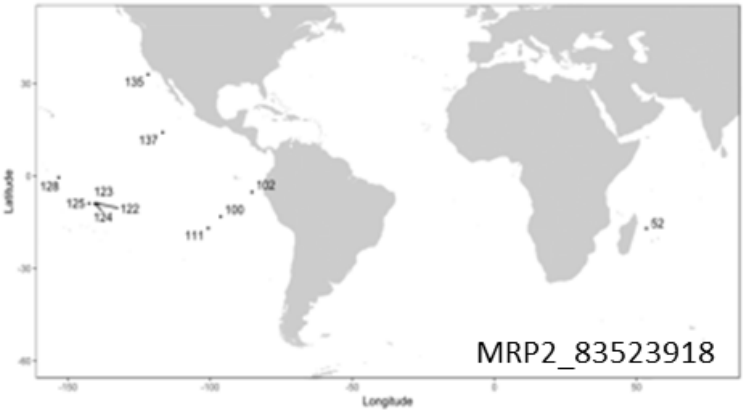

MRP1_89189946
Latitude
Longitude

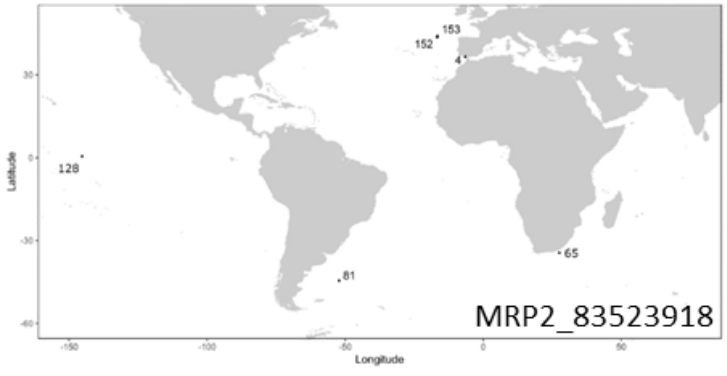

MRP1_29402207
Latitude
Longitude

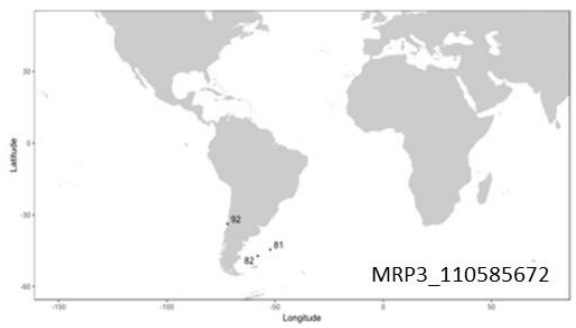

MRP1_80515699
Latitude
Longitude

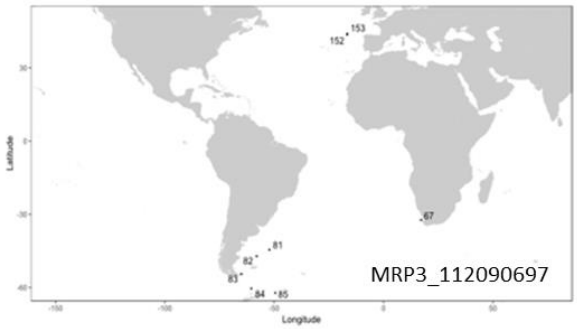

153
152
4
67
81
82
83
84
85
MRP2_53436960

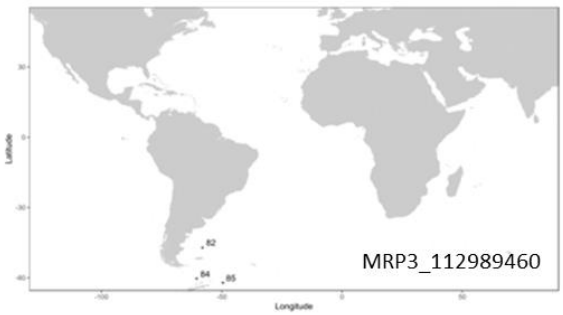

152
142
109
110
102
125
123
100
111
52
64
65
66
80
81
MRP2_97117869
Longitude

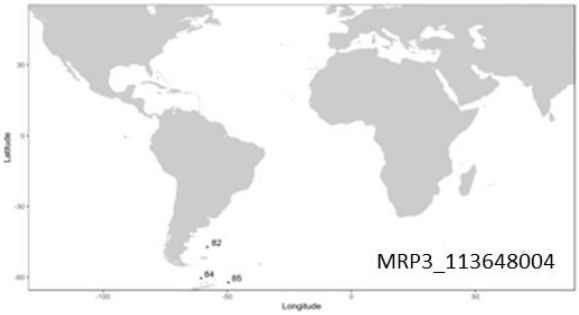

135
137
128
123
125
124
122
102
100
111
52
MRP2_83523918
Latitude
Longitude

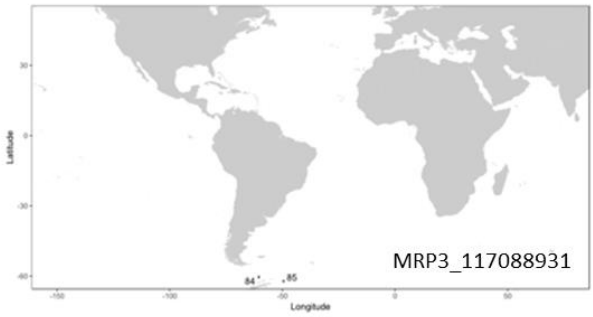

153
152
4
128
65
81
MRP2_83523918
Latitude
Longitude

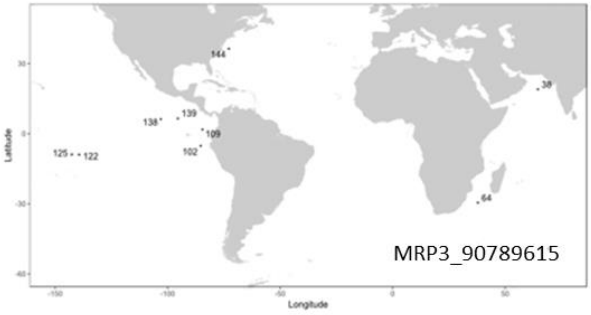

92
81
82
MRP3_110585672
Latitude
Longitude


153
152
67
81
82
83
84
85
MRP3_112090697
Latitude
Longitude


82
84
85
MRP3_112989460
Latitude
Longitude


82
84
85
MRP3_113648004
Latitude
Longitude


84
85
MRP3_117088931
Latitude
Longitude


144
38
139
138
109
125
122
102
64
MRP3_90789615
Latitude
Longitude

References

Adelfi, Maria Grazia, Marco Borra, Remo Sanges, Marina Montresor, Angelo Fontana, and Maria Immacolata Ferrante. 2014. 'Selection and Validation of Reference Genes for QPCR Analysis in the Pennate Diatoms Pseudo-Nitzschia Multistriata and P. Arenysensis'. *Journal of Experimental Marine Biology and Ecology* 451 (February): 74–81. https://doi.org/10.1016/j.jembe.2013.11.003.

Alberts, Bruce. 2015. *Molecular Biology of the Cell*. Sixth edition. New York, NY: Garland Science, Taylor and Francis Group.

Alexander, Harriet, Bethany D. Jenkins, Tatiana A. Rynearson, and Sonya T. Dyhrman. 2015. 'Metatranscriptome Analyses Indicate Resource Partitioning between Diatoms in the Field'. *Proceedings of the National Academy of Sciences* 112 (17): E2182–90. https://doi.org/10.1073/pnas.1421993112.

Amato, A, W Kooistra, J Levialdighiron, D Mann, T Proschold, and M Montresor. 2007. 'Reproductive Isolation among Sympatric Cryptic Species in Marine Diatoms'. *Protist* 158 (2): 193–207. https://doi.org/10.1016/j.protis.2006.10.001.

Amato, Alberto, and Marina Montresor. 2008. 'Morphology, Phylogeny, and Sexual Cycle of Pseudo-Nitzschia Mannii Sp. Nov. (Bacillariophyceae): A Pseudo-Cryptic Species within the P. Pseudodelicatissima Complex'. *Phycologia* 47 (5): 487–97. https://doi.org/10.2216/07-92.1.

Amato, Alberto, Valeria Sabatino, Göran M. Nylund, Johanna Bergkvist, Swaraj Basu, Mats X. Andersson, Remo Sanges, et al. 2018. 'Grazer-Induced Transcriptomic and Metabolomic Response of the Chain-Forming Diatom Skeletonema Marinoi'. *The ISME Journal* 12 (6): 1594–1604. https://doi.org/10.1038/s41396-018-0094-0.

Andersen, Claus Lindbjerg, Jens Ledet Jensen, and Torben Falck Ørntoft. 2004. 'Normalization of Real-Time Quantitative Reverse Transcription-PCR Data: A Model-Based Variance Estimation Approach to Identify Genes Suited for Normalization, Applied to Bladder and Colon Cancer Data Sets'. *Cancer Research* 64 (15): 5245–50. https://doi.org/10.1158/0008-5472.CAN-04-0496.

Apt, Kirk E, AR Grossman, and PG Kroth-Pancic. 1996. 'Stable Nuclear Transformation of the DiatomPhaeodactylum Tricornutum'. *Molecular and General Genetics MGG* 252 (5): 572–579.

Armbrust, E. V. 2004. 'The Genome of the Diatom Thalassiosira Pseudonana: Ecology, Evolution, and Metabolism'. *Science* 306 (5693): 79–86. https://doi.org/10.1126/science.1101156.

Armbrust, E. Virginia. 2009. 'The Life of Diatoms in the World's Oceans'. *Nature* 459 (7244): 185–92. https://doi.org/10.1038/nature08057.

Bachtrog, Doris, Judith E. Mank, Catherine L. Peichel, Mark Kirkpatrick, Sarah P. Otto, Tia-Lynn Ashman, Matthew W. Hahn, et al. 2014. 'Sex Determination: Why So Many Ways of Doing It?' *PLoS Biology* 12 (7): e1001899. https://doi.org/10.1371/journal.pbio.1001899.

Basu, Swaraj, Shrikant Patil, Daniel Mapleson, Monia Teresa Russo, Laura Vitale, Cristina Fevola, Florian Maumus, et al. 2017. 'Finding a Partner in the Ocean: Molecular and Evolutionary Bases of the Response to Sexual Cues in a Planktonic Diatom'. *New Phytologist* 215 (1): 140–56. https://doi.org/10.1111/nph.14557.

Bates, S. S., C J. Bird, A. S. W. de Freitas, R. Foxall, M. Gilgan, L. A. Hanic, G. R. Johnson, et al. 1989. 'Pennate Diatom *Nitzschia Pungens* as the Primary Source of Domoic Acid, a Toxin in Shellfish from Eastern Prince Edward Island, Canada'. *Canadian Journal of Fisheries and Aquatic Sciences* 46 (7): 1203–15. https://doi.org/10.1139/f89-156.

Bates, Stephen S. n.d. ‘Pseudo-Nitzschia, Nitzschia, and Domoic Acid_ New Research since 2011’, 41.

Benton, M. J., and D. A. T. Harper. 2009. *Introduction to Paleobiology and the Fossil Record*. Chichester, UK ; Hoboken, NJ: Wiley-Blackwell.

Bowler, Chris, Andrew E. Allen, Jonathan H. Badger, Jane Grimwood, Kamel Jabbari, Alan Kuo, Uma Maheswari, et al. 2008. ‘The Phaeodactylum Genome Reveals the Evolutionary History of Diatom Genomes’. *Nature* 456 (7219): 239–44. https://doi.org/10.1038/nature07410.

Bromke, Mariusz A., Jamal S. Sabir, Fahad A. Alfassi, Nahid H. Hajarah, Saleh A. Kabli, Abdulrahman L. Al-Malki, Matt P. Ashworth, Michaël Méret, Robert K. Jansen, and Lothar Willmitzer. 2015. ‘Metabolomic Profiling of 13 Diatom Cultures and Their Adaptation to Nitrate-Limited Growth Conditions’. Edited by Adrianna Ianora. *PLOS ONE* 10 (10): e0138965. https://doi.org/10.1371/journal.pone.0138965.

Brunson, John K, Shaun M K McKinnie, Jonathan R Chekan, John P McCrow, Zachary D Miles, Erin M Bertrand, Vincent A Bielinski, et al. 2018. ‘Biosynthesis of the Neurotoxin Domoic Acid in a Bloom-Forming Diatom’, 4.

Chen, Chih Yu. 2004. ‘Biosynthesis of Di-(2-Ethylhexyl) Phthalate (DEHP) and Di-n-Butyl Phthalate (DBP) from Red Alga—Bangia Atropurpurea’. *Water Research* 38 (4): 1014–18. https://doi.org/10.1016/j.watres.2003.11.029.

Chepurnov, Victor A., David G. Mann, Wim Vyverman, Koen Sabbe, and Daniel B. Danielidis. 2002. ‘SEXUAL REPRODUCTION, MATING SYSTEM, AND PROTOPLAST DYNAMICS OF SEMINAVIS (BACILLARIOPHYCEAE)1’. *Journal of Phycology* 38 (5): 1004–19. https://doi.org/10.1046/j.1529-8817.2002.t01-1-01233.x.

Cho, Ji-Young. 2012. ‘Antifouling Activity of Giffinisterone B and Oleamide Isolated from a Filamentous Bacterium Leucothrix Mucor Culture against Ulva Pertusa’. *Korean Journal of Fisheries and Aquatic Sciences* 45 (1): 30–34.

Coelho, S. M., D. Scornet, S. Rousvoal, N. T. Peters, L. Dartevelle, A. F. Peters, and J. M. Cock. 2012. ‘Ectocarpus: A Model Organism for the Brown Algae’. *Cold Spring Harbor Protocols* 2012 (2): pdb.emo065821-pdb.emo065821. https://doi.org/10.1101/pdb.emo065821.

Coleman, Annette W., Lothar Jaenicke, and Richard C. Starr. 2001. ‘GENETICS AND SEXUAL BEHAVIOR OF THE PHEROMONE PRODUCER CHLAMYDOMONAS ALLENSWORTHII (CHLOROPHYCEAE)’. *Journal of Phycology* 37 (2): 345–49. https://doi.org/10.1046/j.1529-8817.2001.037002345.x.

Covington, Brett C., John A. McLean, and Brian O. Bachmann. 2017. ‘Comparative Mass Spectrometry-Based Metabolomics Strategies for the Investigation of Microbial Secondary Metabolites’. *Natural Product Reports* 34 (1): 6–24. https://doi.org/10.1039/C6NP00048G.

Daboussi, Fayza, Sophie Leduc, Alan Maréchal, Gwendoline Dubois, Valérie Guyot, Christophe Perez-Michaut, Alberto Amato, et al. 2014. ‘Genome Engineering Empowers the Diatom Phaeodactylum Tricornutum for Biotechnology’. *Nature Communications* 5 (1). https://doi.org/10.1038/ncomms4831.

D’Alelio, Domenico, Maurizio Ribera d’Alcalà, Laurent Dubroca, Diana Sarn, Adriana Zingone, and Marina Montresor. 2010. ‘The Time for Sex: A Biennial Life Cycle in a Marine Planktonic Diatom’. *Limnology and Oceanography* 55 (1): 106–14. https://doi.org/10.4319/lo.2010.55.1.0106.

De Riso, Valentina, Raffaella Raniello, Florian Maumus, Alessandra Rogato, Chris Bowler, and Angela Falciatore. 2009. ‘Gene Silencing in the Marine Diatom Phaeodactylum Tricornutum’. *Nucleic Acids Research* 37 (14): e96–e96. https://doi.org/10.1093/nar/gkp448.

Diner, Rachel E., Vincent A. Bielinski, Christopher L. Dupont, Andrew E. Allen, and Philip D. Weyman. 2016. 'Refinement of the Diatom Episome Maintenance Sequence and Improvement of Conjugation-Based DNA Delivery Methods'. *Frontiers in Bioengineering and Biotechnology* 4 (August). https://doi.org/10.3389/fbioe.2016.00065.

Dodd, Antony N., Jörg Kudla, and Dale Sanders. 2010. 'The Language of Calcium Signaling'. *Annual Review of Plant Biology* 61 (1): 593–620. https://doi.org/10.1146/annurev-arplant-070109-104628.

Dunahay, Terri G, Eric E Jarvis, and Paul G Roessler. 1995. 'Genetic Transformation of the Diatoms Cyclotella Cryptica and Navicula Saprophila'. *Journal of Phycology* 31 (6): 1004–1012.

Falkowski, P. G. 1998. 'Biogeochemical Controls and Feedbacks on Ocean Primary Production'. *Science* 281 (5374): 200–206. https://doi.org/10.1126/science.281.5374.200.

Falkowski, Paul G., and Matthew J. Oliver. 2007. 'Mix and Match: How Climate Selects Phytoplankton'. *Nature Reviews Microbiology* 5 (10): 813–19. https://doi.org/10.1038/nrmicro1751.

Field, C. B. 1998. 'Primary Production of the Biosphere: Integrating Terrestrial and Oceanic Components'. *Science* 281 (5374): 237–40. https://doi.org/10.1126/science.281.5374.237.

Forstermann, U., and W. C. Sessa. 2012. 'Nitric Oxide Synthases: Regulation and Function'. *European Heart Journal* 33 (7): 829–37. https://doi.org/10.1093/eurheartj/ehr304.

Frenkel, Johannes, Wim Vyverman, and Georg Pohnert. 2014. 'Pheromone Signaling during Sexual Reproduction in Algae'. *The Plant Journal* 79 (4): 632–44. https://doi.org/10.1111/tpj.12496.

Georg Pohnert, Georg Pohnert Boland, and Boland Boland. 2002. 'The Oxylipin Chemistry of Attraction and Defense in Brown Algae and Diatoms'. *Natural Product Reports* 19 (1): 108–22. https://doi.org/10.1039/a806888g.

Getachew, Paulos, Mehader Getachew, Jin Joo, Yoo Seong Choi, Dong Soo Hwang, and Yong-Ki Hong. 2016. 'The Slip Agents Oleamide and Erucamide Reduce Biofouling by Marine Benthic Organisms (Diatoms, Biofilms and Abalones)'. *Toxicology and Environmental Health Sciences* 8 (5): 341–48. https://doi.org/10.1007/s13530-016-0295-8.

Gillard, Jeroen, Johannes Frenkel, Valerie Devos, Koen Sabbe, Carsten Paul, Martin Rempt, Dirk Inzé, Georg Pohnert, Marnik Vuylsteke, and Wim Vyverman. 2013. 'Metabolomics Enables the Structure Elucidation of a Diatom Sex Pheromone'. *Angewandte Chemie International Edition* 52 (3): 854–57. https://doi.org/10.1002/anie.201208175.

Gilroy, S., N. D. Read, and A. J. Trewavas. 1990. 'Elevation of Cytoplasmic Calcium by Caged Calcium or Caged Inositol Trisphosphate Initiates Stomatal Closure'. *Nature* 346 (August): 769.

Gong, Weida, Jamie Browne, Nathan Hall, David Schruth, Hans Paerl, and Adrian Marchetti. 2017. 'Molecular Insights into a Dinoflagellate Bloom'. *The ISME Journal* 11 (2): 439–52. https://doi.org/10.1038/ismej.2016.129.

Guillard Robert R. L. n.d. *The Culture of Marine Invertebrate Animals*.

Handy, Rachel L.C., H.L. Harb, P. Wallace, Z. Gaffen, K.J. Whitehead, and P.K. Moore. 1996. 'Inhibition of Nitric Oxide Synthase by 1-(2-Trifluoromethylphenyl) Imidazole (TRIM) in Vitro: Antinociceptive and Cardiovascular Effects'. *British Journal of Pharmacology* 119 (2): 423–31. https://doi.org/10.1111/j.1476-5381.1996.tb16003.x.

Holtermann, Karie E., Stephen S. Bates, Vera L. Trainer, Anthony Odell, and E. Virginia Armbrust. 2010. 'MASS SEXUAL REPRODUCTION IN THE

TOXIGENIC DIATOMS PSEUDO-NITZSCHIA AUSTRALIS AND P. PUNGENS (BACILLARIOPHYCEAE) ON THE WASHINGTON COAST, USA1: PSEUDO-NITZSCHIA AUXOSPORULATION IN SITU'. *Journal of Phycology* 46 (1): 41–52. https://doi.org/10.1111/j.1529-8817.2009.00792.x.

Hopes, Amanda, Vladimir Nekrasov, Sophien Kamoun, and Thomas Mock. 2016. 'Editing of the Urease Gene by CRISPR-Cas in the Diatom Thalassiosira Pseudonana'. *Plant Methods* 12 (1). https://doi.org/10.1186/s13007-016-0148-0.

Hurwitz, Bonnie L., and Matthew B. Sullivan. 2013. 'The Pacific Ocean Virome (POV): A Marine Viral Metagenomic Dataset and Associated Protein Clusters for Quantitative Viral Ecology'. Edited by Fabiano Thompson. *PLoS ONE* 8 (2): e57355. https://doi.org/10.1371/journal.pone.0057355.

Huysman, Marie JJ, Cindy Martens, Klaas Vandepoele, Jeroen Gillard, Edda Rayko, Marc Heijde, Chris Bowler, and Dirk Inzé. 2010. 'RGeseeanrchome-Wide Analysis of the Diatom Cell Cycle Unveils a Novel Type of Cyclins Involved in Environmental Signaling', 19.

Ianora, A., M. Boersma, R. Casotti, A. Fontana, J. Harder, F. Hoffmann, H. Pavia, P. Potin, S. A. Poulet, and G. Toth. 2006. 'New Trends in Marine Chemical Ecology'. *Estuaries and Coasts* 29 (4): 531–51. https://doi.org/10.1007/BF02784281.

Ianora, Adrianna, Antonio Miralto, Serge A. Poulet, Ylenia Carotenuto, Isabella Buttino, Giovanna Romano, Raffaella Casotti, et al. 2004. 'Aldehyde Suppression of Copepod Recruitment in Blooms of a Ubiquitous Planktonic Diatom'. *Nature* 429 (6990): 403–7. https://doi.org/10.1038/nature02526.

Jolliffe, Ian. 2011. 'Principal Component Analysis'. In *International Encyclopedia of Statistical Science*, 1094–1096. Springer.

Kaczmarska, Irena, Aloisie Poulíčková, Shinya Sato, Mark B. Edlund, Masahiko Idei, Tsuyoshi Watanabe, and David G. Mann. 2013. 'Proposals for a Terminology for Diatom Sexual Reproduction, Auxospores and Resting Stages'. *Diatom Research* 28 (3): 263–94. https://doi.org/10.1080/0269249X.2013.791344.

Kang, Ji-Young, Issa Bangoura, Ji-Young Cho, Jin Joo, Yoo Seong Choi, Dong Soo Hwang, and Yong-Ki Hong. 2016. 'Antifouling Effects of the Periostracum on Algal Spore Settlement in the Mussel Mytilus Edulis'. *Fisheries and Aquatic Sciences* 19 (1): 7.

Karas, Bogumil J., Rachel E. Diner, Stephane C. Lefebvre, Jeff McQuaid, Alex P.R. Phillips, Chari M. Noddings, John K. Brunson, et al. 2015. 'Designer Diatom Episomes Delivered by Bacterial Conjugation'. *Nature Communications* 6 (1). https://doi.org/10.1038/ncomms7925.

Kazamia, Elena, Robert Sutak, Javier Paz-Yepes, Richard G. Dorrell, Fabio Rocha Jimenez Vieira, Jan Mach, Joe Morrissey, et al. 2018. 'Endocytosis-Mediated Siderophore Uptake as a Strategy for Fe Acquisition in Diatoms'. *Science Advances* 4 (5): eaar4536. https://doi.org/10.1126/sciadv.aar4536.

Keeling, Patrick J., Fabien Burki, Heather M. Wilcox, Bassem Allam, Eric E. Allen, Linda A. Amaral-Zettler, E. Virginia Armbrust, et al. 2014. 'The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): Illuminating the Functional Diversity of Eukaryotic Life in the Oceans through Transcriptome Sequencing'. Edited by Roland G. Roberts. *PLoS Biology* 12 (6): e1001889. https://doi.org/10.1371/journal.pbio.1001889.

Kobe, Bostjan, and Andrey V Kajava. n.d. 'The Leucine-Rich Repeat as a Protein Recognition Motif', 8.

Kroth, Peter G., Atle M. Bones, Fayza Daboussi, Maria I. Ferrante, Marianne Jaubert, Misha Kolot, Marianne Nymark, et al. 2018. 'Genome Editing in Diatoms: Achievements and Goals'. *Plant Cell Reports* 37 (10): 1401–8. https://doi.org/10.1007/s00299-018-2334-1.

Kroth, PG. 2007. ‘Genetic Transformation; a Tool to Study Protein Targeting in Diatoms, Chap. 17’. *Methods in Molecular Biology, 2nd Edn., Totowa, NJ, USA: Humana Press*.

Kumar, Sudhir, Glen Stecher, and Koichiro Tamura. 2016. ‘MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets’. *Molecular Biology and Evolution* 33 (7): 1870–74. https://doi.org/10.1093/molbev/msw054.

Lelong, Aurélie, Hélène Hégaret, Philippe Soudant, and Stephen S Bates. 2012. ‘Pseudo-Nitzschia (Bacillariophyceae) Species, Domoic Acid and Amnesic Shellfish Poisoning: Revisiting Previous Paradigms’. *Phycologia* 51 (2): 168–216. https://doi.org/10.2216/11-37.1.

Lim, Hong Chang, Suh Nih Tan, Sing Tung Teng, Nina Lundholm, Emma Orive, Helena David, Sonia Quijano-Scheggia, et al. 2018. ‘Phylogeny and Species Delineation in the Marine Diatom *Pseudo-Nitzschia* (Bacillariophyta) Using *Cox1* , LSU, and ITS2 RRNA Genes: A Perspective in Character Evolution’. Edited by M. Wood. *Journal of Phycology* 54 (2): 234–48. https://doi.org/10.1111/jpy.12620.

Lohbeck, K. T., U. Riebesell, and T. B. H. Reusch. 2014. ‘Gene Expression Changes in the Coccolithophore Emiliania Huxleyi after 500 Generations of Selection to Ocean Acidification’. *Proceedings of the Royal Society B: Biological Sciences* 281 (1786): 20140003–20140003. https://doi.org/10.1098/rspb.2014.0003.

Lommer, Markus, Michael Specht, Alexandra-Sophie Roy, Lars Kraemer, Reidar Andreson, Magdalena A Gutowska, Juliane Wolf, et al. 2012. ‘Genome and Low-Iron Response of an Oceanic Diatom Adapted to Chronic Iron Limitation’. *Genome Biology* 13 (7): R66. https://doi.org/10.1186/gb-2012-13-7-r66.

Lundholm, Nina, Ojvind Moestrup, Yuichi Kotaki, Kerstin Hoef-Emden, Chris Scholin, and Peter Miller. 2006. ‘INTER- AND INTRASPECIFIC VARIATION OF THE PSEUDO-NITZSCHIA DELICATISSIMA COMPLEX (BACILLARIOPHYCEAE) ILLUSTRATED BY RRNA PROBES, MORPHOLOGICAL DATA AND PHYLOGENETIC ANALYSES1’. *Journal of Phycology* 42 (2): 464–81. https://doi.org/10.1111/j.1529-8817.2006.00211.x.

Maier, I. 1993. ‘Gamete Orientation and Induction of Gametogenesis by Pheromones in Algae and Plants’. *Plant, Cell & Environment* 16 (8): 891–907.

Maier, Ingo, and Dieter G. Müller. 1982. ‘Antheridium Fine Structure and Spermatozoid Release in Laminaria Digitata (Phaeophyceae)’. *Phycologia* 21 (1): 1–8. https://doi.org/10.2216/i0031-8884-21-1-1.1.

Malviya, Shruti, Eleonora Scalco, Stéphane Audic, Flora Vincent, Alaguraj Veluchamy, Julie Poulain, Patrick Wincker, et al. 2016. ‘Insights into Global Diatom Distribution and Diversity in the World’s Ocean’. *Proceedings of the National Academy of Sciences* 113 (11): E1516–25. https://doi.org/10.1073/pnas.1509523113.

Mann, David G., and Pieter Vanormelingen. 2013. ‘An Inordinate Fondness? The Number, Distributions, and Origins of Diatom Species’. *Journal of Eukaryotic Microbiology* 60 (4): 414–20. https://doi.org/10.1111/jeu.12047.

Marchetti, A., D. M. Schruth, C. A. Durkin, M. S. Parker, R. B. Kodner, C. T. Berthiaume, R. Morales, A. E. Allen, and E. V. Armbrust. 2012. ‘Comparative Metatranscriptomics Identifies Molecular Bases for the Physiological Responses of Phytoplankton to Varying Iron Availability’. *Proceedings of the National Academy of Sciences* 109 (6): E317–25. https://doi.org/10.1073/pnas.1118408109.

Mattiello, Teresa, Gabriella Fiore, Euan R. Brown, Marco d’Ischia, and Anna Palumbo. 2010. ‘Nitric Oxide Mediates the Glutamate-Dependent Pathway for Neurotransmission in *Sepia Officinalis* Chromatophore Organs’. *Journal of*

Biological Chemistry 285 (31): 24154–63. https://doi.org/10.1074/jbc.M109.083428.

Mock, Thomas, Robert P. Otillar, Jan Strauss, Mark McMullan, Pirita Paajanen, Jeremy Schmutz, Asaf Salamov, et al. 2017. 'Evolutionary Genomics of the Cold-Adapted Diatom Fragilariopsis Cylindrus'. *Nature* 541 (7638): 536–40. https://doi.org/10.1038/nature20803.

Moeys, Sara, Johannes Frenkel, Christine Lembke, Jeroen T. F. Gillard, Valerie Devos, Koen Van den Berge, Barbara Bouillon, et al. 2016. 'A Sex-Inducing Pheromone Triggers Cell Cycle Arrest and Mate Attraction in the Diatom Seminavis Robusta'. *Scientific Reports* 6 (1). https://doi.org/10.1038/srep19252.

Montresor, Marina, Laura Vitale, Domenico D'Alelio, and Maria Immacolata Ferrante. 2016. 'Sex in Marine Planktonic Diatoms: Insights and Challenges'. *Perspectives in Phycology* 3 (2): 61–75. https://doi.org/10.1127/pip/2016/0045.

Montsant, Anton, Andrew E. Allen, Sacha Coesel, Alessandra De Martino, Angela Falciatore, Manuela Mangogna, Magali Siaut, et al. 2007. 'IDENTIFICATION AND COMPARATIVE GENOMIC ANALYSIS OF SIGNALING AND REGULATORY COMPONENTS IN THE DIATOM *THALASSIOSIRA PSEUDONANA*'. *Journal of Phycology* 43 (3): 585–604. https://doi.org/10.1111/j.1529-8817.2007.00342.x.

Mori, Kenji, and Shin-ichi Takanashi. 1996a. 'Stereoselective Synthesis of Lurlenic Acid and Lurlenol, the Sex Pheromones of the Green Flagellate Chlamydomonas.' *Proceedings of the Japan Academy. Ser. B: Physical and Biological Sciences* 72 (8): 174–77. https://doi.org/10.2183/pjab.72.174.

———. 1996b. 'Synthesis of Lurlene, the Sex Pheromone of the Green Flagellate Chlamydomonas Allensworthii'. *Tetrahedron Letters* 37 (11): 1821–24. https://doi.org/10.1016/0040-4039(96)00127-X.

MΓjller, D G, L Jaenicke, M Donike, and T Akintobi. 1971. 'Sex Attractant in a Brown Alga: Chemical Structure'. *Science (New York, N.Y.)* 171 (3976): 1132. https://doi.org/10.1126/science.171.3976.1132.

Niu, Ying-Fang, Zhi-Kai Yang, Meng-Han Zhang, Cong-Cong Zhu, Wei-Dong Yang, Jie-Sheng Liu, and Hong-Ye Li. 2012. 'Transformation of Diatom Phaeodactylum Tricornutum by Electroporation and Establishment of Inducible Selection Marker'. *BioTechniques*, June. https://doi.org/10.2144/000113881.

Nymark, Marianne, Amit Kumar Sharma, Torfinn Sparstad, Atle M. Bones, and Per Winge. 2016. 'A CRISPR/Cas9 System Adapted for Gene Editing in Marine Algae'. *Scientific Reports* 6 (1). https://doi.org/10.1038/srep24951.

Orsini, L, G Procaccini, D Sarno, and M Montresor. 2004. 'Multiple RDNA ITS-Types within the Diatom Pseudo-Nitzschia Delicatissima (Bacillariophyceae) and Their Relative Abundances across a Spring Bloom in the Gulf of Naples'. *Marine Ecology Progress Series* 271: 87–98. https://doi.org/10.3354/meps271087.

Orsini, Luisa, Diana Sarno, Gabriele Procaccini, Roberto Poletti, Jens Dahlmann, and Marina Montresor. 2002. 'Toxic Pseudo-Nitzschia Multistriata (Bacillariophyceae) from the Gulf of Naples: Morphology, Toxin Analysis and Phylogenetic Relationships with Other Pseudo-Nitzschia Species'. *European Journal of Phycology* 37 (2): 247–57. https://doi.org/10.1017/S0967026202003608.

Patil, Shrikant, Sara Moeys, Peter von Dassow, Marie J. J. Huysman, Daniel Mapleson, Lieven De Veylder, Remo Sanges, Wim Vyverman, Marina Montresor, and Maria Immacolata Ferrante. 2015. 'Identification of the Meiotic Toolkit in Diatoms and Exploration of Meiosis-Specific SPO11 and RAD51 Homologs in the Sexual Species Pseudo-Nitzschia Multistriata and Seminavis Robusta'. *BMC Genomics* 16 (1). https://doi.org/10.1186/s12864-015-1983-5.

Pesant, Stéphane, Fabrice Not, Marc Picheral, Stefanie Kandels-Lewis, Noan Le Bescot, Gabriel Gorsky, Daniele Iudicone, et al. 2015. 'Open Science Resources for the Discovery and Analysis of Tara Oceans Data'. *Scientific Data* 2 (May): 150023. https://doi.org/10.1038/sdata.2015.23.

Pierson, Leland S., and Elizabeth A. Pierson. 2010. 'Metabolism and Function of Phenazines in Bacteria: Impacts on the Behavior of Bacteria in the Environment and Biotechnological Processes'. *Applied Microbiology and Biotechnology* 86 (6): 1659–70. https://doi.org/10.1007/s00253-010-2509-3.

Pratiwi, Alberta Rika, Dahrul Syah, Linawati Hardjito, Lily Maria Goretti Panggabean, and Maggy Thenawidjaja Suhartono. 2009. 'Fatty Acid Synthesis by Indonesian Marine Diatom, Chaetoceros Gracilis'. *HAYATI Journal of Biosciences* 16 (4): 151–56. https://doi.org/10.4308/hjb.16.4.151.

Ramesh, M, S Malik, and J Logsdon. 2005. 'A Phylogenomic Inventory of Meiotic GenesEvidence for Sex in Giardia and an Early Eukaryotic Origin of Meiosis'. *Current Biology* 15 (2): 185–91. https://doi.org/10.1016/S0960-9822(05)00028-X.

Rayko, Edda, Florian Maumus, Uma Maheswari, Kamel Jabbari, and Chris Bowler. 2010. 'Transcription Factor Families Inferred from Genome Sequences of Photosynthetic Stramenopiles'. *New Phytologist* 188 (1): 52–66. https://doi.org/10.1111/j.1469-8137.2010.03371.x.

Rines, Jeb, Pl Donaghay, Mm Dekshenieks, Jm Sullivan, and Ms Twardowski. 2002. 'Thin Layers and Camouflage: Hidden Pseudo-Nitzschia Spp. (Bacillariophyceae) Populations in a Fjord in the San Juan Islands, Washington, USA'. *Marine Ecology Progress Series* 225: 123–37. https://doi.org/10.3354/meps225123.

Round, Frank Eric, Richard M Crawford, and David G Mann. 1990. *Diatoms: Biology and Morphology of the Genera*. Cambridge university press.

Ruggiero, Maria Valeria, Domenico D'Alelio, Maria Immacolata Ferrante, Mariano Santoro, Laura Vitale, Gabriele Procaccini, and Marina Montresor. 2018. 'Clonal Expansion behind a Marine Diatom Bloom'. *The ISME Journal* 12 (2): 463–72. https://doi.org/10.1038/ismej.2017.181.

Russo, Monia T., Laura Vitale, Laura Entrambasaguas, Konstantinos Anestis, Neri Fattorini, Filomena Romano, Carmen Minucci, et al. 2018. 'MRP3 Is a Sex Determining Gene in the Diatom Pseudo-Nitzschia Multistriata'. *Nature Communications* 9 (1). https://doi.org/10.1038/s41467-018-07496-0.

Sabatino, Valeria, Monia Teresa Russo, Shrikant Patil, Giuliana d'Ippolito, Angelo Fontana, and Maria Immacolata Ferrante. 2015. 'Establishment of Genetic Transformation in the Sexually Reproducing Diatoms Pseudo-Nitzschia Multistriata and Pseudo-Nitzschia Arenysensis and Inheritance of the Transgene'. *Marine Biotechnology* 17 (4): 452–62. https://doi.org/10.1007/s10126-015-9633-0.

Sabelis, Maurice W., and Marcel Dicke. 1987. 'How Plants Obtain Predatory Mites as Bodyguards'. *Netherlands Journal of Zoology* 38 (2): 148–65. https://doi.org/10.1163/156854288X00111.

Saitou, Naruya, and Masatoshi Nei. 1987. 'The Neighbor-Joining Method: A New Method for Reconstructing Phylogenetic Trees.' *Molecular Biology and Evolution* 4 (4): 406–425.

Sarno, Diana, Adriana Zingone, and Marina Montresor. 2010. 'A Massive and Simultaneous Sex Event of Two Pseudo-Nitzschia Species'. *Deep Sea Research Part II: Topical Studies in Oceanography* 57 (3–4): 248–55. https://doi.org/10.1016/j.dsr2.2009.09.012.

Sato, Shinya, Gordon Beakes, Masahiko Idei, Tamotsu Nagumo, and David G. Mann. 2011. 'Novel Sex Cells and Evidence for Sex Pheromones in Diatoms'. Edited

by Jean-Marc A. Lobaccaro. *PLoS ONE* 6 (10): e26923. https://doi.org/10.1371/journal.pone.0026923.
Scalco, Eleonora, Alberto Amato, Maria Immacolata Ferrante, and Marina Montresor. 2016. 'The Sexual Phase of the Diatom Pseudo-Nitzschia Multistriata: Cytological and Time-Lapse Cinematography Characterization'. *Protoplasma* 253 (6): 1421–31. https://doi.org/10.1007/s00709-015-0891-5.
Scalco, Eleonora, Krzysztof Stec, Daniele Iudicone, Maria Immacolata Ferrante, and Marina Montresor. 2014. 'The Dynamics of Sexual Phase in the Marine Diatom *Pseudo-Nitzschia Multistriata* (Bacillariophyceae)'. Edited by C. Bowler. *Journal of Phycology* 50 (5): 817–28. https://doi.org/10.1111/jpy.12225.
Sims, Patricia A., David G. Mann, and Linda K. Medlin. 2006. 'Evolution of the Diatoms: Insights from Fossil, Biological and Molecular Data'. *Phycologia* 45 (4): 361–402. https://doi.org/10.2216/05-22.1.
Smetacek, V. S. 1985. 'Role of Sinking in Diatom Life-History Cycles: Ecological, Evolutionary and Geological Significance'. *Marine Biology* 84 (3): 239–51. https://doi.org/10.1007/BF00392493.
Smetacek, Victor. 1999. 'Diatoms and the Ocean Carbon Cycle'. *Protist* 150 (1): 25–32. https://doi.org/10.1016/S1434-4610(99)70006-4.
Smith, Colin A, Grace O'Maille, Elizabeth J Want, Chuan Qin, Sunia A Trauger, Theodore R Brandon, Darlene E Custodio, Ruben Abagyan, and Gary Siuzdak. 2005. 'METLIN: A Metabolite Mass Spectral Database'. *Therapeutic Drug Monitoring* 27 (6): 747–751.
Smith, Colin A., Elizabeth J. Want, Grace O'Maille, Ruben Abagyan, and Gary Siuzdak. 2006. 'XCMS: Processing Mass Spectrometry Data for Metabolite Profiling Using Nonlinear Peak Alignment, Matching, and Identification'. *Analytical Chemistry* 78 (3): 779–87. https://doi.org/10.1021/ac051437y.
Sorhannus, Ulf. 2007. 'A Nuclear-Encoded Small-Subunit Ribosomal RNA Timescale for Diatom Evolution'. *Marine Micropaleontology* 65 (1–2): 1–12. https://doi.org/10.1016/j.marmicro.2007.05.002.
Starr, R. C., F. J. Marner, and L. Jaenicke. 1995. 'Chemoattraction of Male Gametes by a Pheromone Produced by Female Gametes of Chlamydomonas.' *Proceedings of the National Academy of Sciences* 92 (2): 641–45. https://doi.org/10.1073/pnas.92.2.641.
Sunagawa, S., L. P. Coelho, S. Chaffron, J. R. Kultima, K. Labadie, G. Salazar, B. Djahanschiri, et al. 2015. 'Structure and Function of the Global Ocean Microbiome'. *Science* 348 (6237): 1261359–1261359. https://doi.org/10.1126/science.1261359.
Tanaka, Tsuyoshi, Yoshiaki Maeda, Alaguraj Veluchamy, Michihiro Tanaka, Heni Abida, Eric Maréchal, Chris Bowler, et al. 2015. 'Oil Accumulation by the Oleaginous Diatom *Fistulifera Solaris* as Revealed by the Genome and Transcriptome'. *The Plant Cell Online* 27 (1): 162–76. https://doi.org/10.1105/tpc.114.135194.
Tara Oceans Coordinators, Quentin Carradec, Eric Pelletier, Corinne Da Silva, Adriana Alberti, Yoann Seeleuthner, Romain Blanc-Mathieu, et al. 2018. 'A Global Ocean Atlas of Eukaryotic Genes'. *Nature Communications* 9 (1). https://doi.org/10.1038/s41467-017-02342-1.
Tautenhahn, Ralf, Christoph Bottcher, and Steffen Neumann. 2008. 'Highly Sensitive Feature Detection for High Resolution LC/MS'. *BMC Bioinformatics* 9 (1): 504. https://doi.org/10.1186/1471-2105-9-504.
Tonon, Thierry, David Harvey, Tony R Larson, and Ian A Graham. 2002. 'Long Chain Polyunsaturated Fatty Acid Production and Partitioning to Triacylglycerols in Four Microalgae'. *Phytochemistry* 61 (1): 15–24.

Torii, Keiko U. 2004. ‘Leucine-Rich Repeat Receptor Kinases in Plants: Structure, Function, and Signal Transduction Pathways’. In *International Review of Cytology*, 234:1–46. Elsevier. https://doi.org/10.1016/S0074-7696(04)34001-5.

Traller, Jesse C., Shawn J. Cokus, David A. Lopez, Olga Gaidarenko, Sarah R. Smith, John P. McCrow, Sean D. Gallaher, et al. 2016. ‘Genome and Methylome of the Oleaginous Diatom Cyclotella Cryptica Reveal Genetic Flexibility toward a High Lipid Phenotype’. *Biotechnology for Biofuels* 9 (1). https://doi.org/10.1186/s13068-016-0670-3.

Treguer, P., D. M. Nelson, A. J. Van Bennekom, D. J. DeMaster, A. Leynaert, and B. Queguiner. 1995. ‘The Silica Balance in the World Ocean: A Reestimate’. *Science* 268 (5209): 375–79. https://doi.org/10.1126/science.268.5209.375.

Trygg, Johan, and Torbjörn Lundstedt. 2007. ‘Chemometrics Techniques for Metabonomics’. In *The Handbook of Metabonomics and Metabolomics*, 171–99. Elsevier. https://doi.org/10.1016/B978-044452841-4/50007-2.

TSUBO, YOSHIHIRO. 1961. ‘Chemotaxis and Sexual Behavior in Chlamydomonas’. *The Journal of Protozoology* 8 (2): 114–121.

Vardi, Assaf, Fabio Formiggini, Raffaella Casotti, Alessandra De Martino, François Ribalet, Antonio Miralto, and Chris Bowler. 2006. ‘A Stress Surveillance System Based on Calcium and Nitric Oxide in Marine Diatoms’. Edited by Jeffrey Dangl. *PLoS Biology* 4 (3): e60. https://doi.org/10.1371/journal.pbio.0040060.

Vargas, C. de, S. Audic, N. Henry, J. Decelle, F. Mahe, R. Logares, E. Lara, et al. 2015. ‘Eukaryotic Plankton Diversity in the Sunlit Ocean’. *Science* 348 (6237): 1261605–1261605. https://doi.org/10.1126/science.1261605.

Villar, Emilie, Thomas Vannier, Caroline Vernette, Magali Lescot, Aurelien Alexandre, Paul Bachelerie, Thomas Rosnet, Eric Pelletier, Shinichi Sunagawa, and Pascal Hingamp. 2018. ‘The Ocean Gene Atlas: Exploring the Biogeography of Plankton Genes Online’, February. https://doi.org/10.1101/271882.

Watson, A. J., D. C. E. Bakker, A. J. Ridgwell, P. W. Boyd, and C. S. Law. 2000. ‘Effect of Iron Supply on Southern Ocean CO2 Uptake and Implications for Glacial Atmospheric CO2’. *Nature* 407 (6805): 730–33. https://doi.org/10.1038/35037561.

Williamson, Shannon J., Douglas B. Rusch, Shibu Yooseph, Aaron L. Halpern, Karla B. Heidelberg, John I. Glass, Cynthia Andrews-Pfannkoch, et al. 2008. ‘The Sorcerer II Global Ocean Sampling Expedition: Metagenomic Characterization of Viruses within Aquatic Microbial Samples’. Edited by Neil Hall. *PLoS ONE* 3 (1): e1456. https://doi.org/10.1371/journal.pone.0001456.

Wood, EJ. 1983. ‘Molecular Cloning. A Laboratory Manual by T Maniatis, EF Fritsch and J Sambrook. Pp 545. Cold Spring Harbor Laboratory, New York. 1982. $48 ISBN 0-87969-136-0’. *Biochemical Education* 11 (2): 82–82.

Yoon, Hwan Su, Jeremiah D. Hackett, Claudia Ciniglia, Gabriele Pinto, and Debashish Bhattacharya. 2004. ‘A Molecular Timeline for the Origin of Photosynthetic Eukaryotes’. *Molecular Biology and Evolution* 21 (5): 809–18. https://doi.org/10.1093/molbev/msh075.

Zhang, Chunye, and Hanhua Hu. 2014. ‘High-Efficiency Nuclear Transformation of the Diatom Phaeodactylum Tricornutum by Electroporation’. *Marine Genomics* 16 (August): 63–66. https://doi.org/10.1016/j.margen.2013.10.003.